



Estimation Bayésienne de l'abondance par "removal sampling" en présence de variabilité du taux d'échantillonnage : application aux tiques *Ixodes ricinus* en quête d'hôtes

Severine Bord

► To cite this version:

Severine Bord. Estimation Bayésienne de l'abondance par "removal sampling" en présence de variabilité du taux d'échantillonnage : application aux tiques *Ixodes ricinus* en quête d'hôtes. Sciences agricoles. Université Blaise Pascal - Clermont-Ferrand II, 2014. Français. NNT : 2014CLF22463 . tel-01166340

HAL Id: tel-01166340

<https://theses.hal.science/tel-01166340>

Submitted on 22 Jun 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**ECOLE DOCTORALE SCIENCES DE LA VIE, SANTÉ, AGRONOMIE,
ENVIRONNEMENT**
N° d'ordre 632

THÈSE

Présentée à l'INRA de Clermont-Ferrand – Theix

pour l'obtention du grade de
DOCTEUR D'UNIVERSITÉ
Spécialité : Écologie

Soutenue le 17 juin 2014
par
Séverine Bord

**ESTIMATION BAYÉSIENNE DE L'ABONDANCE PAR
« REMOVAL SAMPLING » EN PRÉSENCE DE VARIABILITÉ
DU TAUX D'ÉCHANTILLONNAGE :
APPLICATION AUX TIQUES *Ixodes ricinus* EN QUÊTE D'HÔTES**

Jury

M.	HERVÉ MONOD	INRA, Jouy-en-Josas	Président
M.	CHRISTOPHE ABRAHAM	INRA, Montpellier	Rapporteur
Mme	KAREN MCCOY	CNRS, Montpellier	Rapporteur
M.	CHRISTIAN AMBLARD	CNRS, Clermont-Ferrand	Examineur
Mme	GWENAËL VOURC'H	INRA, Clermont-Ferrand	Co-Directrice de Thèse
M.	PIERRE DRUILHET	Université Blaise Pascal, Clermont-Ferrand	Directeur de Thèse

Table des matières

1	Introduction	5
1.1	Indicateurs d'abondance et taux de détection	6
1.2	Les méthodes d'échantillonnage pour mesurer les abondances : principes et limites	9
1.2.1	Comptage unique	9
1.2.2	« Distance Sampling »	9
1.2.3	Capture Marquage Recapture	10
1.2.4	« Removal Sampling »	10
1.3	L'estimation de l'abondance des tiques	11
1.3.1	Intérêt épidémiologique des tiques	12
1.3.2	Cycle d'activité des tiques <i>I. ricinus</i>	13
1.3.3	Choix du « removal sampling » pour estimer l'abondance des tiques	14
1.4	Modélisation de données issues du « removal sampling »	17
1.4.1	Les différents modèles probabilistes	17
1.4.1.1	Modélisation pour une unité d'observation	18
1.4.1.2	Modélisation pour plusieurs unités d'observation	19
1.4.2	Les méthodes d'estimation des paramètres	22
1.4.2.1	Méthodes fréquentistes	22
1.4.2.2	Méthodes Bayésiennes	24
1.5	Objectifs et plan de la thèse	27
2	Les techniques d'échantillonnage des tiques : principes et limites	29
3	Conditions environnementales, taux d'échantillonnage et indicateurs d'abondance	41
4	Modélisation Bayésienne hiérarchique de la taille de la population et du taux d'échantillonnage	65

5	Influence de l'<i>a priori</i> sur les estimateurs Bayésiens	77
5.1	Schéma de simulations et critère de comparaison des estimateurs . . .	79
5.2	Résultats des simulations	81
5.2.1	Comportement des estimateurs dans le cas non convergent (cas 5)	81
5.2.2	Comportement des estimateurs dans le cas convergent (cas 4 et 6)	84
5.3	Article en préparation	88
6	Discussion générale et perspectives	101
6.1	« Removal Sampling » : avantages et limites de la méthode adaptée aux tiques	102
6.2	Valeurs estimées du taux d'échantillonnage des tiques <i>I. ricinus</i> . . .	103
6.3	Importance de la variabilité du taux dans l'estimation de l'abondance	105
6.4	Choix des <i>a priori</i>	107
6.5	Estimation du risque acarologique, comment répartir la capacité d'échantillonnage ?	107
6.6	Conclusion	109
7	Annexes	111
7.1	Pré-étude de CMR	111
7.1.1	Peu ou pas de recapture	111
7.1.2	Augmentation du nombre de recaptures au fur et à mesure des échantillonnages	111
7.1.3	Décroissance puis augmentation du nombre de recaptures au fur et à mesure des échantillonnages	112
7.2	Vraisemblance et vraisemblance profilée	115

Introduction

Les changements globaux, tels que la modification du climat, les changements d'usage des terres et des milieux aquatiques bouleversent les interactions entre les espèces d'hôtes (Guégan et al., 2004; Tylianakis et al., 2008; Begon et al., 2009) en modifiant la répartition des espèces (Marsot et al., 2013). Ces interactions perturbent l'équilibre de l'écosystème¹ en modifiant la répartition des ressources, en bouleversant les comportements et en transformant les systèmes écologiques. Ces changements engendrent de nouveaux contacts entre des hôtes et des agents pathogènes dont les rapprochements nouveaux, multiples et répétés offrent des conditions favorables à l'émergence de nouvelles maladies (Morse, 1995; Jones et al., 2008).

Dans ce contexte, l'étude des variations spatiales et temporelles de l'abondance des espèces, *i.e.* du nombre d'individus d'une même espèce dans une unité de surface à un temps donné (Williams et al., 2002), est une étape primordiale pour comprendre les mécanismes qui génèrent les modifications des écosystèmes et des dynamiques de population, pour définir et proposer des actions de régulation, et enfin pour comprendre les effets de ces actions afin d'anticiper sur de prochaines mesures (Yoccoz et al., 2001; Coreau et al., 2009). Cette étape est d'autant plus importante qu'elle concerne une espèce de vecteur² impliquée dans un cycle épidémiologique (*e.g.* moustiques, tiques, rongeurs); en effet pour estimer le risque de maladies induit par un vecteur, il est important de pouvoir déterminer les conditions qui favorisent son existence et/ou son maintien. Si les conditions favorables à la présence du vecteur engendrent une mauvaise détection des individus par rapport aux autres conditions alors le risque lié à ce vecteur sera mal estimé. Il est donc nécessaire de disposer d'estimations d'abondance pertinentes et de qualité, cela repose sur l'effet combiné de méthodes d'échantillonnage adaptées et de la mise en œuvre de méthodes d'estimations performantes.

Au cours de la section 1.1, nous définissons l'abondance et présentons la no-

1. ensemble formé par une association ou une communauté entre des êtres vivants et leur environnement abiotique.

2. organisme qui intervient dans la transmission d'un agent pathogène, transmission qui peut être inter-humaine ou de l'animal à l'homme.

tion d'indicateur d'abondance. Dans la section 1.2, nous présentons les principales méthodes d'échantillonnage utilisées pour collecter des données d'abondance, leurs avantages et leurs limites. Dans la section 1.3, nous présentons la problématique liée aux estimations d'abondance des tiques. La section 1.4 se focalise sur les méthodes de « removal sampling » et présente les modèles envisageables et les méthodes pour estimer la taille de la population et le taux d'échantillonnage. Enfin, la section 1.5 pose la problématique et les objectifs appliqués et théoriques de la thèse.

1.1 Indicateurs d'abondance et taux de détection

L'étude de l'abondance d'une espèce repose très souvent sur l'usage d'indicateurs d'abondance. En effet, bien que d'un intérêt majeur, l'abondance est une quantité qui ne peut être observée directement car tous les individus présents sur une unité d'observation ne sont pas forcément observables (Yoccoz et al., 2001; Williams et al., 2002). Ce défaut d'observation est un phénomène courant en écologie qui conduit à un taux de détection inférieur à 1 (Williams et al., 2002; Schmidt, 2005; Royle and Dorazio, 2008); il peut être lié au comportement de l'espèce (domaine d'activité de l'espèce trop important pour pouvoir être échantillonné dans sa globalité), à la technique d'échantillonnage (plus ou moins efficace) ou encore aux conditions d'échantillonnage (expérience de l'observateur, climat, végétation, ...).

Si l'on considère n unités d'observation et une détection parfaite des individus, *i.e.* un taux de détection égal à 1, alors le nombre d'individus détectés sur chacune des n unités d'observation, sachant le nombre d'individus présents, est complètement déterminé et sa variabilité est uniquement liée à la variabilité spatiale et/ou temporelle de la population présente sur chacune des n unités d'observation (cf. Encart 1.1, cas $\tau = 1$).

Par contre, si le taux de détection est inférieur à 1, alors le nombre d'individus détectés sur chacune des n unités d'observation sachant le nombre d'individus présents, devient une variable aléatoire qui génère une variabilité supplémentaire (cf. Encart 1.1, cas $\tau = \tau_0$).

De plus, dans la majorité des cas, le taux varie d'une unité d'observation à l'autre (Thompson et al., 1998; Royle et al., 2005; Royle and Dorazio, 2008; McClintock et al., 2010). Dans ces cas de figure, la variabilité du nombre d'individus détectés sera d'autant plus forte (cf. Encart 1.1, cas $\tau < 1$).

Si on ne tient pas compte de la variabilité de la détection entre unités d'observation, les conclusions concernant la population faites à partir des n sous-échantillons (Royle and Dorazio, 2008) seront incertaines jusqu'à être erronées (Conroy, 1996; Anderson, 2003; Jennelle et al., 2007). En effet, les écarts observés entre les n valeurs de l'indicateur d'abondance basés sur le nombre de captures des n échantillons ne peuvent pas être attribués uniquement à un écart d'abondance mais aussi à un écart

du taux de détection (Johnson, 1995; MacKenzie and Kendall, 2002; Pollock et al., 2002; Williams et al., 2002). Il est donc nécessaire dans un premier temps d'estimer le taux de détection et sa variabilité en fonction des conditions d'échantillonnage pour pouvoir répondre à la question : *Y a-t-il plus d'individus détectés dans A par rapport à B parce que l'abondance de A est supérieure à celle de B ou parce que les individus de A sont mieux détectés que ceux de B ?* L'objectif à terme est d'estimer une abondance « ajustée » par la valeur du taux d'échantillonnage pour chacune des unités d'observation pour pouvoir réaliser des comparaisons spatiales et/ou temporelles des abondances, inférer sur des dynamiques de population ou mettre en évidence des facteurs de risque liés aux abondances.

Encart 1.1: Décomposition de la variabilité de l'estimateur de la taille de la population N_0 .

Soit x la quantité d'individus observés sur une unité d'observation et τ le taux d'échantillonnage, on a alors :

$$\begin{aligned}(x|N_0, \tau) &\sim B(N_0, \tau) \\ \Rightarrow E(x|N_0, \tau) &= N_0\tau \\ \Rightarrow V(x|N_0, \tau) &= N_0\tau(1 - \tau)\end{aligned}$$

L'estimateur \hat{N}_0 de la taille de la population N_0 est $\hat{N}_0 = \frac{x}{\tau}$.

- si $\tau = 1$:

$$\begin{aligned}E(\hat{N}_0|N_0, \tau) &= N_0 \\ V(\hat{N}_0|N_0, \tau) &= 0\end{aligned}$$

- si $\tau = \tau_0$ avec $\tau_0 < 1$:

$$\begin{aligned}E(\hat{N}_0|N_0, \tau) &= N_0\tau_0 \\ V(\hat{N}_0|N_0, \tau) &= N_0 \frac{1-\tau_0}{\tau_0} \quad \text{si} \quad \begin{aligned} &\tau_0 \rightarrow 0, V(\hat{N}_0) \rightarrow \infty \\ &\tau_0 \rightarrow 1, V(\hat{N}_0) \rightarrow 0 \end{aligned}\end{aligned}$$

- si $\tau < 1, E(\tau) = \tau_0$ et $V(\tau) = \sigma_\tau^2$:

$$\begin{aligned}E(\hat{N}_0|N_0, \tau) &= N_0\tau_0 \\ V(\hat{N}_0|N_0, \tau) &= N_0 \left(\frac{(N_0-1)\sigma_\tau^2}{\tau_0^2} + \frac{1-\tau_0}{\tau_0} \right)\end{aligned}$$

Lorsque τ varie, la variance de l'estimateur de la taille de la population $V(\hat{N}_0)$ est liée à :

- la taille de la population : N_0
- la valeur même de τ : $\frac{1-\tau_0}{\tau_0}$
- la variabilité de τ : $\frac{(N_0-1)\sigma_\tau^2}{\tau_0^2}$

1.2 Les méthodes d'échantillonnage pour mesurer les abondances : principes et limites

Pour estimer des abondances plusieurs méthodes d'échantillonnage sont envisageables. Nous détaillons ici les 4 principales méthodes.

1.2.1 Comptage unique

L'échantillonnage par « comptage unique » ou « point count survey » en anglais consiste à compter un nombre d'individus détectés ou capturés par unité d'observation à un instant donné. Ce type d'échantillonnage est couramment utilisé en écologie pour estimer des abondances d'espèces (Rosenstock et al., 2002; Williams et al., 2002; Elphick, 2008). Les données de comptage obtenues fournissent simplement un indicateur d'abondance car la totalité des individus présents dans chaque unité d'observation n'est pas détectée ou capturée; on parle de comptages bruts ou « raw count » en anglais. Ces données de comptage issues des différentes unités d'observation sont ensuite comparées afin de conclure sur des écarts d'abondance de population entre les unités d'observation. Le taux de détection (ou de capture) est supposé constant dans le temps et dans l'espace ce qui autorise la comparaison directe des données de comptage entre elles sans avoir à estimer l'abondance.

Cependant, cette méthode d'échantillonnage est souvent critiquée (Burnham, 1981; Johnson, 1995; Anderson, 2001, 2003) car elle repose sur l'hypothèse forte de taux de détection constant alors que cette hypothèse n'est que peu vérifiée. Dans ce cas, la variabilité du taux de détection (ou de capture) s'ajoute à la variabilité naturelle des données de comptage liée à la variabilité d'abondance, participant ainsi à une augmentation de la sur-dispersion des données (Lindén and Mäntyniemi, 2011). Ainsi, les conclusions des modèles qui cherchent à mettre en évidence les facteurs associés à l'abondance seront erronées (Tyre et al., 2003; Gu and Swihart, 2004) : par exemple, si les individus sont moins bien détectés dans un habitat A qui est défavorable à la présence de l'espèce par rapport à un autre habitat B , alors l'effet du type d'habitat B sur l'abondance sera surestimé. À l'inverse, si les individus sont mieux détectés dans A par rapport à B alors l'effet du type d'habitat B sur l'abondance sera sous-estimé voire non significatif bien qu'il existe.

1.2.2 « Distance Sampling »

Les méthodes dites de « Distance Sampling » (que l'on pourrait traduire par « échantillonnage par les distances ») consistent à mesurer la distance entre l'observateur et l'animal détecté au moment de l'observation (Burnham, 1981; Buckland et al., 2005). Ces méthodes reposent sur le fondement théorique selon lequel la probabilité de détection des individus présents sur une unité d'observation est fonction de

la distance à l'observateur. Cette fonction est *a priori* décroissante, *i.e.* plus l'individu est loin de l'observateur et plus la probabilité de détection sera faible. L'abondance est alors estimée à partir des distances. Cette méthode peut être envisagée de 2 façons différentes : échantillonnages par observation à partir de tronçons ou de points. Ces méthodes impliquent que l'espèce observée soit visible à distance et que les distances soient bien estimées.

1.2.3 Capture Marquage Recapture

Les méthodes de Capture Marquage Recapture (CMR) sont couramment utilisées pour étudier les abondances (Lebreton et al., 1992). Elles consistent à capturer un échantillon aléatoire d'individus de taille n_1 parmi une population. Les individus capturés sont marqués puis relâchés dans la population. Un second échantillonnage de taille n_2 est alors réalisé dans cette même population (Williams et al., 2002). On note m_2 , le nombre d'individus marqués capturés au second échantillonnage. Les hypothèses de cette méthode sont que le marquage est permanent (ne peut être perdu) et que le comportement des individus marqués est comparable à celui des individus non marqués.

L'estimation \hat{N} de la taille de la population est alors égale à :

$$\hat{N} = \frac{n_1 \cdot n_2}{m_2}$$

Les méthodes CMR sont généralement coûteuses en termes de main d'œuvre (Pollock et al., 2002; Dodd and Dorazio, 2004). De plus, pour certaines espèces, ce type d'échantillonnage ne peut pas être envisagé car les étapes de capture, de marquage et de réinsertion dans la population modifient la survie ou le comportement des individus marqués. Par exemple, ces modifications se traduisent par une probabilité de capture différente entre les individus marqués et non marqués car les individus marqués mémorisent la première capture (Crowcroft and Jeffreers, 1961; Gilbert et al., 2001).

1.2.4 « Removal Sampling »

L'échantillonnage par prélèvements successifs ou « removal sampling », « depletion survey » en anglais consiste à répéter des échantillonnages sur une même unité d'observation. Les individus sont capturés successivement et sans remise avec la même probabilité parmi la population (Williams et al., 2002). L'intervalle de temps entre deux prélèvements successifs est généralement court pour préserver l'hypothèse majeure de population fermée *i.e.* pas d'immigration, d'émigration, de naissance ou de mort (MacKenzie and Royle, 2005). Les quantités successives d'individus capturés sont ensuite modélisées pour estimer la probabilité de détection et la taille de la

population présente sur l'unité d'observation. Tout comme les CMR, les méthodes par « removal sampling » sont coûteuses en termes de main d'œuvre (Dodd and Dorazio, 2004) mais présentent l'avantage d'être une alternative aux méthodes de CMR lorsque celles-ci ne sont pas envisageables (modification du comportement des individus marqués par exemple).

1.3 L'estimation de l'abondance des tiques

Dans la section 1.2.1, nous avons vu que les estimations d'abondance basées sur des comptages uniques reposent sur l'hypothèse forte de taux d'échantillonnage constant. En écologie, cette hypothèse est très peu vérifiée ce qui peut conduire à des conclusions erronées sur les abondances : par exemple, la mise en évidence de l'effet significatif d'un facteur sur l'abondance alors qu'il influence la détection. Dans ce contexte, nous nous sommes intéressés à un exemple d'application, à savoir, l'estimation de l'abondance des tiques *Ixodes ricinus*.

L'abondance des tiques *I. ricinus* d'une unité d'observation est classiquement estimée par le nombre de tiques capturées lors d'un seul échantillonnage. Chaque unité d'observation est décrite selon des facteurs tels que le type de végétation, la nature du milieu à proximité, l'exposition, Les quantités de captures sont alors comparées afin de mettre en évidence les facteurs significativement associés aux écarts d'abondance (Boyard et al., 2007, 2008). Cependant, de nombreux auteurs évoquent une variabilité probable du taux de détection des tiques par la méthode de capture classiquement utilisée, à savoir la méthode du drapeau (méthode décrite dans le chapitre 2) (Milne, 1943; Gray et al., 1978; Gray, 1985; Ostfeld et al., 1996; Schulze et al., 1997). Par conséquent, le risque acarologique³, auquel nous sommes exposés, est susceptible d'être mal estimé. Pour autant ce taux de détection n'a été que peu étudié et rarement estimé. A notre connaissance, seules 2 études (Talleklint-Eisen and Lane, 2000; Daniels et al., 2000) ont estimés ce taux d'échantillonnage à moins de 10%. Cette estimation du taux est obtenue pour d'autres espèces qu'*I. ricinus* qui sont susceptibles d'avoir un comportement différent (vis-à-vis de la réactivité au passage du drapeau ou de part la hauteur de quête sur la végétation), à savoir, *Ixodes scapularis* pour Daniels et al. (2000) et *Ixodes pacificus* pour Talleklint-Eisen and Lane (2000).

Dans les paragraphes suivants, nous allons présenter l'intérêt épidémiologique des tiques. Nous détaillerons ensuite le cycle d'activité des tiques au sein d'un stade de développement et le type de population de tiques échantillonnées par la méthode du drapeau. Enfin, nous présenterons notre approche par « removal sampling » pour l'estimation de l'abondance des tiques.

3. densité de tiques infectées en quête d'hôte.

1.3.1 Intérêt épidémiologique des tiques

Présentes dans la plupart des régions du monde, les tiques sont des acariens qui parasitent les vertébrés dont les mammifères et l'homme (Parola and Raoult, 2001). Impliquées dans la transmission de nombreux agents pathogènes (bactéries, virus, ...), les tiques engendrent des problèmes sanitaires importants en médecine humaine et vétérinaire (Parola, 2004).

Ixodes ricinus, l'espèce de tique la plus répandue en France transmet notamment l'agent de la borréliose de Lyme, maladie vectorielle la plus prévalente dans les zones tempérées de l'hémisphère nord (Matuschka and Spielman, 1986; Parola and Raoult, 2001). Le cycle de vie d'*I. ricinus* se décompose en 3 stades actifs : larves, nymphes et adultes (cf. Fig. 1.1). Le passage d'un stade au suivant se fait par la prise d'un repas sanguin sur des hôtes. Les larves sont ubiquistes pour ses hôtes *i.e.* elles se gorgent sur la plupart des vertébrés allant du micro-mammifère à l'homme en passant par les oiseaux. Les nymphes préfèrent des hôtes de petite ou moyenne taille (micro-mammifères, oiseaux, léporidés, ...) et les adultes privilégient les grands mammifères (cervidés, bovins, ...). La présence de tiques sur ces différents types d'hôtes contribue, notamment, à disperser les tiques dans différents milieux.

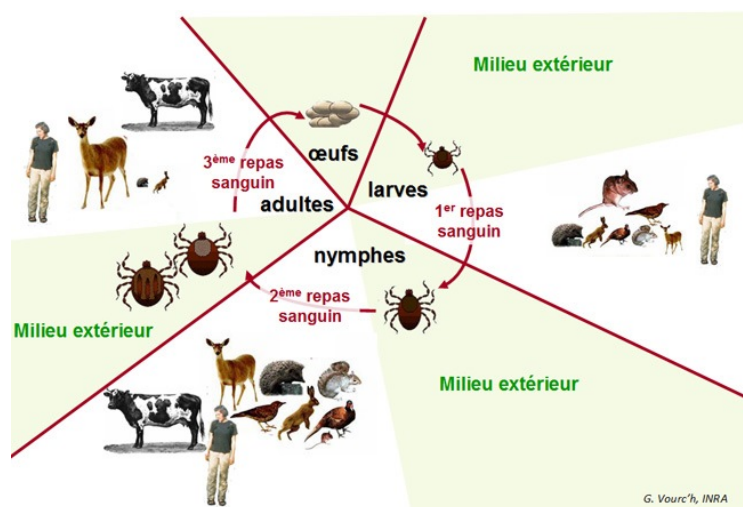


Fig. 1.1 : Cycle de vie des tiques *Ixodes ricinus* (Source : G. Vourc'h)

Au cours des repas sanguins, les tiques porteuses d'un agent pathogène sont susceptibles de contaminer leur hôte, ou de contracter un nouvel agent pathogène, dans le cas où l'hôte est déjà porteur. Parmi les 3 stades actifs, le stade nymphal présente un intérêt majeur car les nymphes ont la capacité de transmettre l'agent de la maladie de Lyme, contrairement aux larves car la transmission transovarienne de cette bactérie est peu probable (Schoeler and Lane, 1993; Nefedova et al., 2004). De plus, même si les nymphes ont un taux d'infection inférieur à celui des adultes

(en moyenne 18,6% pour les adultes contre 10,1% pour les nymphes (Rauter and Hartung, 2005)), elles présentent une densité beaucoup plus importante car le taux de mortalité entre les stades est élevé (Randolph and Craine, 1995). Enfin, de part leur très petite taille, les nymphes sont moins facilement détectables que les tiques adultes, ce qui leur laisse le temps de transmettre un agent pathogène à un hôte par une morsure plus longue (Compte rendu de la 16^{ème} conférence en thérapeutique anti-infectieuse).

Les nymphes représentent donc un risque majeur de transmission de pathogènes. Pour évaluer les risques associés à la transmission d'agents pathogènes par les tiques, il est donc nécessaire d'estimer leur abondance.

1.3.2 Cycle d'activité des tiques *I. ricinus*

La période durant laquelle les tiques sont à la recherche d'un hôte est communément appelée « période d'activité » en opposition aux phases de digestion, de développement et de mue qui se déroulent après le repas de sang et qui sont des phases non actives. Deux pics d'activité des tiques *I. ricinus*, plus ou moins marqués selon les années, sont décrits dans la littérature : un pic d'avril à juillet puis un second de septembre à octobre (Lees and Milne, 1951; Gray, 1982, 1985; Walker, 2001; Randolph et al., 2002; Randolph, 2004). L'intensité de ces deux pics d'activité fluctue selon les conditions climatiques et micro-climatiques⁴ (Gray, 1982; Randolph et al., 2002). La saisonnalité de la période d'activité des tiques est liée au mécanisme de diapause⁵ qui génère un arrêt momentané du développement en attendant des conditions de survie plus favorables pour trouver un hôte, se nourrir et muer au stade suivant (Belozerov, 1982). Généralement, la mise en activité des tiques est associée une augmentation de la température qui coïncide avec un ensemble de conditions favorables à la mise à l'affût des tiques pour la recherche d'un hôte (Estrada-Peña et al., 2013). La mise à l'affût consiste à grimper sur un support ce qui permet aux tiques de se placer sur le haut de la végétation pour attendre le passage d'un hôte potentiel (cf. Fig. 1.2). Lorsque l'hôte de passage s'approche suffisamment, elles s'y accrochent et recherchent un endroit favorable pour mordre (MacLeod, 1932; Gigon, 1985).

Lorsque les nymphes sont à l'affût sur le haut de la végétation, elles sont soumises à des conditions climatiques (exposition au soleil, vent, ...) qui accélèrent leur dessiccation à laquelle elles sont extrêmement sensibles (Lees, 1946; Sonenshine, 1994). Par conséquent, les tiques actives vont alterner des phases de quête à l'affût et de réhydratation au sol (Lees and Milne, 1951; Perret et al., 2003; Estrada-Peña et al.,

4. le micro environnement du sol et l'interface entre la végétation et le sol constituent un micro-climat

5. phase génétiquement déterminée dans le développement d'un organisme où il diminue l'intensité de ses activités métaboliques.



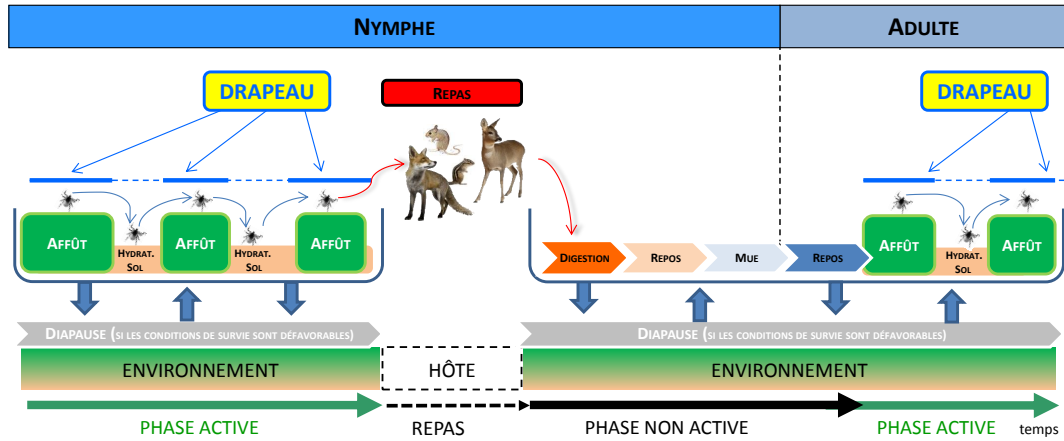
Fig. 1.2 : *Ixodes ricinus* adulte femelle à l'affût (Source : C. Boyard)

2013) (cf. Fig. 1.3). La durée de la quête à l'affût est inversement associée au niveau de dessiccation (Perret et al., 2003; Randolph et al., 2002; Estrada-Pena et al., 2004) et le temps nécessaire à la réhydratation peut être très variable selon les individus (Kahl and Knülle, 1988).

Parmi la population active, on distingue donc deux types de population : la population active à l'affût et la population active au sol. Parmi ces deux populations, seule la population active à l'affût sera accessible à l'échantillonnage par la méthode du drapeau (cf Fig. 1.3). Cette méthode consiste à balayer la végétation d'une unité d'observation à l'aide d'un drap d'un mètre carré (cf. chapitre 2 pour plus de détails sur la méthode du drapeau).

1.3.3 Choix du « removal sampling » pour estimer l'abondance des tiques

Classiquement, l'estimation de l'abondance des tiques *I. ricinus* est basée sur la mise en œuvre des méthodes d'échantillonnage par comptage unique. Cependant, ces méthodes peuvent conduire à de mauvaises estimations de l'abondance dans la mesure où l'hypothèse de taux d'échantillonnage constant par la méthode du drapeau est remise en question dans la littérature (Milne, 1943; Gray et al., 1978; Gray, 1985; Ostfeld et al., 1996; Schulze et al., 1997). Dans ce contexte, il paraît important d'étudier la qualité de l'appareil de mesure *i.e.* estimer le taux d'échantillonnage de la méthode du drapeau et sa variabilité selon les conditions d'échantillonnage des unités d'observation (type de végétation, exposition, saison, ...). Pour cela, nous avons mené une pré-étude par Capture Marquage Recapture (CMR), en mai 2011,



Pas à l'échelle de temps réelle

Fig. 1.3 : Activité d'une tique au cours du stade nymphal et échantillonnage par la méthode du drapeau

pour tester la faisabilité d'une telle méthode sur les tiques (cf. Encart 1.2).

Les principaux résultats de cette pré-étude nous amènent à penser que le marquage a un effet sur le comportement des nymphes marquées. L'application aux tiques de la méthode par CMR nécessite donc une phase de mise au point que nous n'avons pas souhaité réaliser dans le cadre de ce travail de thèse. C'est donc pour cette raison que nous avons privilégié une approche par « removal sampling ».

Cette approche par « removal sampling » consistait à échantillonner une unité d'observation bornée de 1m × 10m (cf. Fig. 1.4) par des échantillonnages successifs réalisés toutes les 2 minutes 30s.

Encart 1.2:

Etude par Capture Marquage Recapture des nymphes *I. ricinus*

Objectif :

Une pré-étude (non publiée) a été menée en mai 2011 pour tester la faisabilité d'une approche par CMR afin d'étudier les propriétés du taux d'échantillonnage des nymphes à l'affût.

Matériel et Méthodes :

Cette étude consistait à capturer 200 nymphes à l'affût (le matin). Ces nymphes étaient ensuite marquées par une poudre fluorescente (RADGLO[®] TP) puis relâchées (à midi) sur une unité d'observation de 10m² délimitée et dont la végétation était homogène. 15 échantillonnages successifs (l'après-midi) étaient réalisés sur cette unité d'observation toutes les 10 minutes. Entre le lâché et le 1^{er} échantillonnage, un intervalle de temps de 2h30 était laissé pour permettre aux nymphes marquées de réintégrer la population et de reprendre leur quête d'un hôte. Enfin, 10 échantillonnages supplémentaires étaient réalisés le lendemain matin. Les nymphes récoltées ont été stockées dans des tubes individuels pour éviter une contamination du marquage. Elles ont ensuite été visualisées à l'aide d'un microscope à filtre UV pour déterminer leur statut vis-à-vis du marquage. Un total de 11 unités d'observation ont été échantillonnées.

Résultats et discussion :

Nous avons noté une variabilité importante entre les sites du nombre de nymphes marquées-recapturées au cours des 15 échantillonnages successifs (de 0 à 38%) et de l'allure des courbes de captures. Les types de courbe étaient les suivants :

- très peu de recapture voire aucune recapture (cf. Fig. 7.1 en annexe) ;
- un nombre de nymphes marquées et recapturées croissant au cours des 1^{ers} échantillonnages (cf. Fig. 7.2 en annexe) ;
- un nombre de nymphes capturées (marquées ou non) décroissant au cours des 1^{ers} échantillonnages (rangs 1 à 4/5), croissant aux rangs suivants (rangs 5/6-7/8) et décroissant à nouveau lors des derniers échantillonnages (rangs 8 et plus) (cf. Fig. 7.3 en annexe).

Ces résultats laissent supposer que : (i) l'intervalle de temps entre le lâché et le 1^{er} échantillonnage (2h30) était insuffisant pour permettre aux nymphes marquées de reprendre leur quête à l'affût ; (ii) l'intervalle de temps entre les échantillonnages successifs (10min) était trop long pour éviter que les tiques actives au sol ne se mettent à l'affût au fur et à mesure des échantillonnages.



Fig. 1.4 : Exemple d'une unité d'observation de « removal sampling »

1.4 Modélisation de données issues du « removal sampling »

Au paragraphe 1.2, nous avons détaillé les principales méthodes d'échantillonnages permettant de récolter des données utiles à l'estimation d'abondance d'espèces. Les indicateurs d'abondances basés sur une seule observation par site n'étant pas satisfaisants (taux de détection variable généralement), les méthodes par distance sampling et les méthodes par CMR n'étant pas applicables à certaines espèces comme les tiques, nous nous sommes intéressés aux méthodes d'échantillonnage par « removal sampling ».

Dans cette partie, nous allons détailler les différents modèles de probabilité qui peuvent être construits pour décrire des données de RS puis nous présenterons ensuite les deux méthodes d'estimations envisageables.

1.4.1 Les différents modèles probabilistes

On définit une unité d'observation comme une zone échantillonnée par « removal sampling ». Pour cette unité d'observation, on cherche à estimer la taille de la population et le taux d'échantillonnage de la méthode d'observation des individus. La population de l'unité d'observation est considérée comme fermée *i.e.* sans immigration, émigration, naissance ou mort au cours des échantillonnages successifs. Nous présentons ici plusieurs types de modélisation pour des données de removal sampling issues d'une seule unité d'observation puis de plusieurs unités d'observation.

1.4.1.1 Modélisation pour une unité d'observation

Soit $X = (X_1, \dots, X_K)$, la séquence de K captures observées sur une unité d'observation donnée. On pose X_k , le nombre de captures au $k^{\text{ème}}$ échantillonnage, N_k la taille de la population restante après le $k^{\text{ème}}$ échantillonnage avec $N_k = N_{k-1} - X_k$ pour $k \in 1 \dots K$ et τ_k le taux d'échantillonnage au rang k . On considère alors que les X_k suivent une distribution binomiale de paramètres N_{k-1} et τ_k :

$$(X_k | N_{k-1}, \tau_k) \sim \mathcal{B}(N_{k-1}, \tau_k), \text{ où } N_k = N_{k-1} - X_k. \quad [1.1]$$

La probabilité de capture τ_k représente la probabilité pour un individu d'être détecté par la technique d'échantillonnage (observateur, méthode du drapeau, ...) au $k^{\text{ème}}$ échantillonnage. Le taux d'échantillonnage représente la proportion de la population détectée par la technique d'échantillonnage. On suppose que les individus de la population N_{k-1} sont capturés indépendamment les uns des autres et avec la même probabilité τ_k ; le taux d'échantillonnage est alors égal à la probabilité de capture.

Le taux d'échantillonnage peut être envisagé comme plus ou moins constant selon les conditions liées aux caractéristiques de l'espèce étudiée, de l'unité d'observation ou des propriétés connues ou non de la technique d'échantillonnage. Différents modèles peuvent alors être envisagés, allant du modèle le plus simple (taux constant) vers des modèles plus complexes (taux variables).

- Modèle à taux constant :

Le taux d'échantillonnage est constant au cours des K échantillonnages successifs (Moran, 1951; Zippin, 1958; Dodd and Dorazio, 2004; Royle, 2004a,b; Royle and Dorazio, 2006) :

$$\tau_k = \tau, k = 1, \dots, K. \quad [1.2]$$

- Modèle à taux variable :

Le taux d'échantillonnage varie au cours des K échantillonnages successifs :

$$\tau_k \neq \tau_{k'}, \text{ pour } k \neq k'. \quad [1.3]$$

Cependant, les τ_k ne sont pas identifiables car en considérant une seule unité d'observation, nous ne disposons que d'une seule observation X_k par rang d'échantillonnage k . Ce problème d'identifiabilité peut être contourné en définissant une structure temporelle particulière sur le taux d'échantillonnage ; par exemple, le taux d'échantillonnage de rang k peut être défini comme égal à celui du rang $k - 1$ à un facteur aléatoire près, indépendant de k noté ϵ (Dauphin et al., 2009; Brun et al., 2011) :

$$\text{logit}(\tau_k) = \text{logit}(\tau_{k-1}) - \epsilon, \text{ où } \text{logit}(\tau) = \ln\left(\frac{\tau}{1-\tau}\right). \quad [1.4]$$

Cette structure temporelle annule le problème d'identifiabilité en réduisant le nombre de paramètres à estimer deux paramètres (τ_1 et ϵ) contre k (τ_1, \dots, τ_K).

1.4.1.2 Modélisation pour plusieurs unités d'observation

On considère simultanément m unités d'observation échantillonnées par « removal sampling » dans des conditions différentes (heure, saison, végétation, observateur, ...). On note N_0^i , la taille de la population sur l'unité d'observation i . On pose $X^i = (X_1^i, \dots, X_K^i)$, la séquence de captures observées sur l'unité d'observation i pour $i = 1, \dots, m$ où K représente le nombre total d'échantillonnages successifs réalisés sur chaque unité d'observation. On note X_k^i , le nombre de captures au $k^{\text{ème}}$ échantillonnage pour l'unité d'observation i , N_k^i la taille de la population restante après le $k^{\text{ème}}$ échantillonnage avec $N_k^i = N_{k-1}^i - X_k^i$ et τ_k^i , le taux d'échantillonnage au rang k pour le site i avec $k = 1, \dots, K$.

On considère que les X_k^i suivent une distribution binomiale de paramètres N_{k-1}^i et τ_k^i :

$$(X_k^i | N_{k-1}^i, \tau_k^i) \sim \mathcal{B}(N_{k-1}^i, \tau_k^i), \text{ où } N_k^i = N_{k-1}^i - X_k^i. \quad [1.5]$$

Ce modèle n'est pas identifiable pour tous les τ_k^i et N_0^i d'où l'importance de définir une structure sur les paramètres. Nous allons détailler ici différentes structures pour les paramètres τ_k^i et N_0^i que l'on cherche à estimer.

Structures envisageables pour τ_k^i :

- Modèle à taux constant : le taux d'échantillonnage est constant entre les unités d'observation i et les rangs d'échantillonnage k :

$$\tau_k^i = \tau, \text{ pour } i = [1, \dots, m] \text{ et } k = [1, \dots, K]. \quad [1.6]$$

- Modèle à structure temporelle : le taux d'échantillonnage est variable selon les rangs d'échantillonnages k mais constant entre les unités d'observation pour une même valeur de k :

$$\begin{aligned} \tau_k^i &\neq \tau_{k'}^i \text{ pour } k \neq k' \text{ et } i = [1, \dots, m] \\ \text{et } \tau_k^i &= \tau_k^j \text{ pour } i \neq j \text{ et } k = [1, \dots, K]. \end{aligned} \quad [1.7]$$

- Modèle à structure spatiale : le taux d'échantillonnage est variable d'une unité d'observation à l'autre mais constant quel que soit le rang de l'échantillonnage (pas de variabilité temporelle) :

$$\begin{aligned} \tau_k^i &\neq \tau_k^j \text{ pour } i \neq j \text{ et } k = [1, \dots, K] \\ \text{et } \tau_k^i &= \tau_{k'}^i \text{ pour } k \neq k' \text{ et } i = [1, \dots, m]. \end{aligned} \quad [1.8]$$

Cette variabilité spatiale du taux peut être liée aux caractéristiques propres à chacune des unités d'observation échantillonnées, qu'elles soient connues ou non, observables ou non. Si ces caractéristiques sont connues et observables, les τ^i sont alors vus comme des effets fixes. Par contre, si les variations des τ^i sont dues à des facteurs non observables, les τ^i sont alors vus comme des effets aléatoires dont la distribution peut être approchée par une loi $\beta(a, b)$ (Wyatt, 2002), ou une loi $N(\mu, \sigma^2)$ pour les $\text{logit}(\tau^i)$ (Rivot et al., 2008; Bohrmann et al., 2012).

- Modèle à structure définie par des covariables :

Le taux d'échantillonnage dépend de covariables liées à l'unité d'observation mais constant quel que soit le rang d'échantillonnage. On note $F^i = (F_1^i, \dots, F_r^i)$ les r covariables associées à l'unité d'observation i . On modélise τ_k^i par :

$$\tau_k^i = g(F^i), \text{ avec } g(.) \text{ la fonction de lien.} \quad [1.9]$$

avec $g(.)$ la fonction de lien

L'effet des covariables sur les variations de τ est classiquement introduit sous la forme d'un modèle linéaire généralisé (Royle, 2004b; Rivot et al., 2008; Dauphin et al., 2009) :

$$\text{logit}(\tau_k^i) = \alpha_0 + \sum_{j=1}^r F_j^i \cdot \alpha_j, \text{ avec } \text{logit}(\tau) = \ln\left(\frac{\tau}{1-\tau}\right).$$

Les $\text{logit}(\tau_k^i)$ sont vus comme des réalisations d'une variable aléatoire dont la distribution est donnée par (Rivot et al., 2008; Dauphin et al., 2009) :

$$\text{logit}(\tau_k^i | \alpha_j, \sigma) \sim \mathcal{N}((\alpha_0 + \sum_{j=1}^r F_j^i \cdot \alpha_j), \sigma^2). \quad [1.10]$$

- Modèle à structure spatiale et temporelle : le taux d'échantillonnage est variable entre les unités d'observation et selon le rang d'échantillonnage :

$$\tau_k^i \neq \tau_{k'}^j \text{ pour } i \neq j \text{ et } k \neq k'. \quad [1.11]$$

Ce modèle n'est pas identifiable. Par exemple, Mantyniemi et al. (2005) contourne ce problème d'identifiabilité en définissant une structure temporelle et spatiale du taux où $\tau_k^i = \mu^i \frac{\eta^i}{\eta^i + k - 1}$.

Structures envisageables pour N_0^i :

- Aucune structure : les tailles des populations des m unités d'observation sont indépendantes et sans aucun lien ni structure.
- Structure temporelle : les tailles des populations des m unités d'observation varient selon le rang d'échantillonnage indépendamment des captures réalisées lors des échantillonnages. On parle de population ouverte. Ces variations au cours des échantillonnages successifs peuvent être liées à des arrivées ou des départs d'individus (immigration, migration, naissances ou morts). Dans le cadre de notre travail, nous avons considéré une population fermée.
- Structure spatiale : les tailles des populations des m unités d'observation sont distribuées selon une même loi donnée. Le plus souvent, il s'agit d'une loi de Poisson dont le paramètre suit une loi Gamma (Wyatt, 2002; Dauphin et al., 2009; Bohrmann et al., 2012).
- Structure définie par des covariables : les tailles des populations des m unités d'observation peuvent dépendre de covariables. On note $F^i = (F_1^i, \dots, F_r^i)$ les r covariables associées à l'unité d'observation i :

$$N_0^i = h(F^i), \text{ avec } h(.) \text{ la fonction de lien.} \quad [1.12]$$

L'effet des covariables sur la taille de la population peut être modélisé par un modèle log-linéaire (Royle, 2004a,b; Royle and Dorazio, 2006).

Par exemple, Royle (2004a) modélise la taille de la population N_0 par une loi de Poisson de paramètre λ et introduit l'effet des covariables F^i sur la variabilité

de N_0^i par :

$$\log(\lambda^i) = \alpha_0 + \sum_{j=1} r\alpha_j \cdot F_j^i.$$

1.4.2 Les méthodes d'estimation des paramètres

Dans la littérature, deux grands types d'approches sont proposés pour estimer la taille de la population N_0 et le taux d'échantillonnage τ à partir de données de captures obtenues par « removal sampling » : les approches fréquentistes et les approches bayésiennes (voir Truong-Van and Prouzet, 2001). Nous présentons ici les principaux éléments.

1.4.2.1 Méthodes fréquentistes

Les moindres carrés

Les premiers développements proposés pour estimer la taille d'une population N_0 et le taux d'échantillonnage τ à partir de données de « removal sampling » ont été établis par Leslie and Davis (1939) et Hayne (1949). Leur méthode consiste à réaliser une régression linéaire entre le nombre de captures obtenues à chaque échantillonnage et la somme cumulée des captures issues des échantillonnages précédents (cf. Fig. 1.5). Le taux d'échantillonnage τ est estimé par la pente de la droite de régression et la taille de la population N_0 par le point d'intersection de la droite de régression avec l'axe des abscisses.

Cette méthode correspond au modèle [1.2] pour lequel le taux d'échantillonnage est constant. Elle ne permet pas de traiter simultanément les données de « removal sampling » issues de plusieurs unités d'observation.

Cette méthode d'estimation présente l'avantage d'être facile à mettre en œuvre. Par contre, elle est connue pour fournir de mauvaises estimations de N_0 en estimant parfois une taille de population inférieure au total des captures (Schnute, 1983). De plus, cette méthode d'estimation ne respecte pas les hypothèses classiques de la régression linéaire, *i.e.* l'homoscédasticité et l'indépendance des erreurs (White et al., 1982).

Le maximum de vraisemblance

Une approche par maximum de vraisemblance peut également être utilisée pour estimer N_0 et τ à partir de données de « removal sampling ». Moran (1951) est le premier à avoir proposé un modèle multinomial pour estimer simultanément le couple (N_0, τ) à partir des propriétés asymptotiques de l'estimateur du maximum de vraisemblance. Zippin (1956, 1958) proposera ensuite une méthode graphique basée sur le maximum de vraisemblance pour estimer les paramètres. Ils montrent

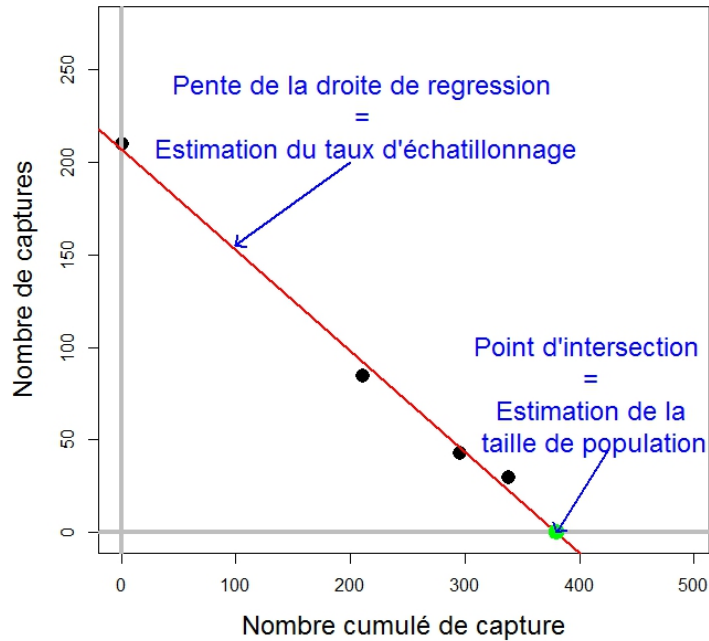


Fig. 1.5 : Régression linéaire

que lorsque le nombre d'échantillonnages successifs est égal à 2, les résultats sont équivalents à ceux obtenus par la méthode de régression (Leslie and Davis, 1939; Hayne, 1949).

Cette approche par maximum de vraisemblance autorise une modélisation plus riche que le modèle à taux constant (modèle [1.2]). Elle permet de traiter les données de plusieurs unités d'observation simultanément en prenant en compte les différentes structures vues précédemment pour le taux d'échantillonnage et la taille de la population (cf. paragraphe 1.4.1.2).

Cependant, l'approche par maximum de vraisemblance dans le cadre de « removal sampling » présente de nombreuses limites :

- l'approche par maximum de vraisemblance conduit régulièrement à des estimations infinies pour N_0 (Carle and Strub, 1978; Schnute, 1983; Bolfarine et al., 1992; Bedrick, 1994). Bedrick (1994) sera le premier à définir les conditions nécessaires et suffisantes à l'existence d'un maximum de vraisemblance et confirme la conjecture proposée par Carle and Strub (1978) permettant de déterminer si une estimation de N_0 par la vraisemblance est finie ou non.
- lorsque le maximum de vraisemblance converge, les estimations privilégiées de N_0 sont des estimations « basses » qui sont le plus souvent très proches de la somme cumulée des captures (Schnute, 1983). Plusieurs auteurs (Schnute,

1983; Gove et al., 1995; Borchers et al., 2002) associent ce « biais conditionnel » de l'estimateur de N_0 à une mauvaise spécification du modèle *i.e.* au fait que l'hypothèse de taux d'échantillonnage constant n'est pas valide.

- les intervalles de confiance des estimateurs sont souvent basés sur l'approximation normale (Seber, 1982; Hirst, 1994), mais cette approximation n'est valable que pour des N_0 grands (Harding et al., 1984). Ces intervalles de confiance sont alors peu fiables et comprennent souvent des valeurs aberrantes (borne inférieure de l'intervalle inférieure à la somme cumulée des captures). Hirst (1994) propose des intervalles de confiance basés sur les rapports de log-vraisemblance profilées⁶. Il démontre par simulations que ces intervalles de confiance sont plus proches de la vraie valeur que ceux basés sur la vraisemblance asymptotique. Cependant, ces approximations asymptotiques ne sont pas valables lorsque le taux d'échantillonnage est faible (Carle and Strub, 1978).

1.4.2.2 Méthodes Bayésiennes

Compte tenu des limites des méthodes par maximum de vraisemblance énoncées au paragraphe précédent de nombreux auteurs se sont tournés vers des approches bayésiennes (Bolfarine et al., 1992; Ellison, 2004; Schwarz and Seber, 1999). Cette tendance est d'autant plus marquée que les méthodes de Monte Carlo et les capacités calculatoires pour estimer les lois *a posteriori* ont été largement développées. Les méthodes Bayésiennes permettent de considérer aisément des modèles probabilistes sophistiqués comme le modèle [1.5] et elles se prêtent facilement à l'introduction d'effets aléatoires par l'implémentation de modèles hiérarchiques (Rivot et al., 2008; Dauphin et al., 2009).

L'estimation des lois *a posteriori* est le résultat de la combinaison entre les connaissances disponibles *a priori* sur les paramètres et l'observation des données (cf. Encart 1.3 pour plus de détails sur les méthodes Bayésiennes).

D'un point de vue pratique, le choix de la loi *a priori* est souvent délicat car l'interprétation de la connaissance *a priori* disponible est le plus souvent insuffisante pour permettre de déterminer une seule et même loi *a priori*. Dans ce cas, des distributions dites « non informatives⁷ » sont généralement utilisées. L'utilisation de ce type de loi repose sur le fait qu'en l'absence de connaissance, on ne souhaite pas influencer les inférences par de l'information externe, l'objectif étant de « laisser les observations parler d'elles-mêmes ». De nombreux travaux, initiés par Bayes and Price (1763), Laplace (1786), Jeffreys (1946) puis Bernardo and Smith (1994) ont tenté de proposer des *a priori* avec lesquels la subjectivité du statisticien interfère au minimum.

6. Vraisemblance profilée de $\hat{\tau} = \arg \max_{\tau} (V|N_0)$: vraisemblance maximale pour chaque valeur potentielle de N_0 (cf. exemple Fig. 7.4 en annexe)

7. à faible contenu informatif

Le choix des distributions *a priori* peut être délicat et conduire à des estimateurs divergents. Cet aspect sera discuté au cours du chapitre 5. Nous allons passer en revue les différentes distributions *a priori* utilisées dans la littérature selon les modèles présentés au paragraphe 1.4.1.2.

Pour le taux d'échantillonnage, la distribution intuitive est une loi uniforme sur $[0, 1]$ (Royle and Dorazio, 2006; Laplace, 1786) qui correspond à une loi $\beta(1, 1)$. Certains auteurs proposent l'*a priori* de Haldane (1932) qui correspond à une loi $\beta(0, 0)$ qui équivaut à une loi uniforme sur le logit ou encore l'*a priori* de Jeffreys (1946) qui correspond à une loi $\beta(1/2, 1/2)$ qui équivaut à une loi uniforme sur $\sin^{-1} \sqrt{\theta}$.

Pour le modèle probabiliste $\tau_k^i = \mu^i \cdot b_k$ et $b_k = \eta^i / (\eta^i + k - 1)$, Mantyniemi et al. (2005) pose une loi *a priori* $\beta(1, 1, 1, 1)$ sur μ^i et $\mathcal{U}[0, 1]$ sur η^i avec i le site et k le rang d'échantillonnage.

Pour le modèle où, τ^i est un effet aléatoire avec $\tau^i \sim \beta(a, b)$, Bohrmann et al. (2012) considère des lois *a priori* vagues $\Gamma(0, 01, 0, 01)$ sur les paramètres a et b .

Pour modéliser une variation temporelle du taux d'échantillonnage (modèle [1.7]), Dauphin et al. (2009) considère un effet aléatoire ϵ sur lequel il pose un *a priori* vague $\epsilon \sim \mathcal{N}(0, 1000)$.

Rivot et al. (2008) considère τ comme effet aléatoire en prenant le logit(τ^i) $\sim \mathcal{N}(\mu, \sigma^2)$ où, μ est distribué selon une loi *a priori* $\mathcal{N}(0, 1000)$ et σ suit une loi *a priori* $\mathcal{U}(0, 1)$. Il modifie ce premier modèle en intégrant des covariables ayant un effet sur le paramètre μ avec $\mu = \alpha + \beta + \gamma$ et où γ est pris comme effet aléatoire de variance σ^2 . Il pose en *a priori* $\alpha \sim \mathcal{N}(0, V_\alpha = 10)$ et $\beta \sim \mathcal{N}(0, V_\beta = 100)$ et une loi uniforme sur $[0, 1]$ pour σ^2 .

Pour modéliser la variabilité spatiale de la population entre site, le modèle le plus souvent utilisé est un modèle de Poisson dont la loi de mélange est une loi $\Gamma(\lambda, k)$ (Wyatt, 2002; Dauphin et al., 2009; Bohrmann et al., 2012). Les *a priori* posés sur λ et sur k sont des $\Gamma(0, 01, 0, 01)$. Rivot et al. (2008) intègre l'effet de covariables sur le paramètre $\log(\theta) = \alpha + \beta + \gamma$ avec γ pris comme effet aléatoire de variance σ^2 . Il pose en *a priori* $\alpha \sim \mathcal{N}(0, V_\alpha = 100)$ et $\beta \sim \mathcal{N}(0, V_\beta = 1000)$ et une loi uniforme sur $[0, 10]$ pour σ^2 .

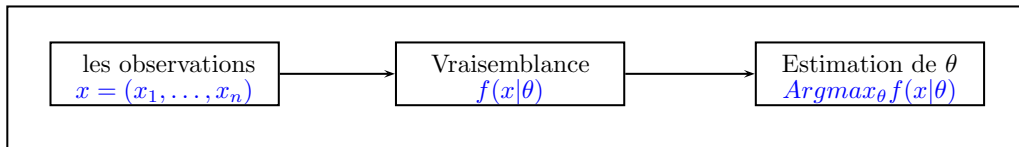
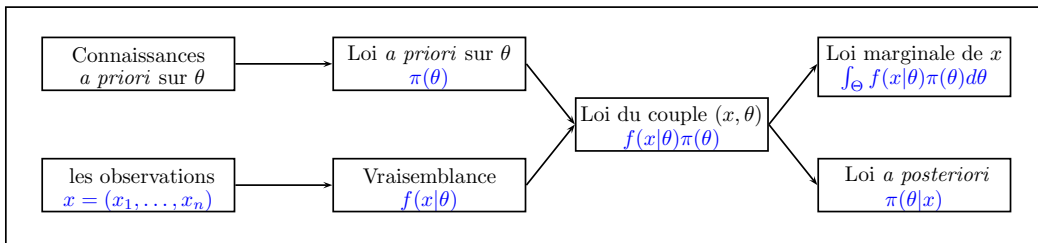
Encart 1.3: **Principe des Méthodes Bayésiennes**

On pose $x = (x_1, \dots, x_i, \dots, x_n)$, l'ensemble des observations et θ un paramètre appartenant à l'ensemble des paramètres Θ . Contrairement à l'approche fréquentiste (cf. Encart 1.4), l'approche bayésienne considère θ comme une variable aléatoire inconnue dont la distribution *a priori* est notée $\pi(\theta)$ (cf. Encart 1.5).

La distribution des observations est décrite par une fonction de densité $f(x|\theta)$, appelée *vraisemblance*. L'objectif est de mettre à jour les connaissances sur la distribution de θ *i.e.* déterminer la loi *a posteriori* notée $\pi(\theta|x)$ à partir des observations x décrites par la vraisemblance $f(x|\theta)$ et la loi *a priori* $\pi(\theta)$. Pour cela, les méthodes bayésiennes reposent sur le théorème de Bayes, on a alors :

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int_{\Theta} f(x|\theta)\pi(\theta)d\theta} \propto f(x|\theta)\pi(\theta).$$

A partir de ces distributions *a posteriori*, on obtient des estimations de θ en considérant la moyenne de $\pi(\theta|x)$ et des régions de crédibilité sont déduits (Gelman et al., 2004; Ellison, 2004). Ils représentent l'intervalle auquel appartient la vraie valeur du paramètre avec une probabilité $1 - \alpha$, où $\alpha = 0,05$ en général.

Encart 1.4: **Principe de l'approche fréquentiste**Encart 1.5: **Principe de l'approche Bayésienne**

1.5 Objectifs et plan de la thèse

La variabilité des données de capture résulte de la variabilité de l'abondance et de la variabilité du taux de capture. L'objectif global de la thèse est d'estimer par une approche bayésienne l'abondance des tiques à partir de données de « removal sampling » en présence de variations du taux d'échantillonnage.

L'objectif appliqué de ce travail de thèse est double : déterminer la méthode d'échantillonnage adaptée pour étudier la variabilité du taux d'échantillonnage en fonction des conditions d'échantillonnage et estimer simultanément pour chaque unité d'observation l'abondance des tiques et le taux d'échantillonnage.

Au cours du chapitre 2, nous détaillons les différentes méthodes d'échantillonnages des tiques, leurs avantages et leurs limites. Au chapitre 3, nous étudions l'effet des facteurs environnementaux sur le taux d'échantillonnage d'une part et le lien entre ces facteurs environnementaux et différents indicateurs d'abondance d'autre part. Nous discutons notamment des conséquences de la variabilité du taux d'échantillonnage sur la qualité de l'estimation de l'abondance. Au chapitre 4, nous développerons un modèle bayésien hiérarchique avec effet aléatoire pour estimer simultanément le taux d'échantillonnage et l'abondance de chaque unité d'observation.

L'objectif théorique est d'étudier les propriétés de l'outil d'estimation bayésienne à savoir la stabilité des estimateurs bayésiens, dans le cadre de données de « removal sampling » et en l'absence de connaissance des distributions *a priori* sur les paramètres. Au cours du chapitre 5, nous étudierons les choix des distributions *a priori* dans le cadre bayésien. Enfin, au chapitre 6, nous discutons des résultats et des perspectives envisageables à notre travail.

Les techniques d'échantillonnage des tiques : principes et limites

Cette partie représente une contribution au livre « *Tiques et maladies à tiques en France* », Ed. K.McCoy et N.Boulanger. Ce livre fait la synthèse de l'état des connaissances actuelles des tiques et des maladies qui leurs sont associées. Il est destiné à un public francophone de scientifiques, décideurs, étudiants, professionnels de la santé. Il est en cours de mise en forme.

L'objectif de cette contribution est de faire l'inventaire des principales techniques d'échantillonnage des tiques.

Dans une première partie, nous présentons les principales techniques d'échantillonnage des tiques dans la végétation et sur les hôtes. Dans une seconde partie, nous présentons les avantages et les inconvénients de chacune des techniques en fonction des caractéristiques des espèces : recherche d'hôtes active (chasse) ou passive (à l'affût), mobilité, . . . Enfin, dans une troisième partie, nous abordons la question des facteurs qui peuvent influencer l'efficacité d'une méthode d'échantillonnage.

Livre « Tiques et maladies à tiques en France »

Ed. K.McCoy et N.Boulanger

Les méthodes d'échantillonnage et fiabilité (S. Bord¹, A. Agoulon²)

¹Unité d'épidémiologie Animale-UR346, Centre INRA de Theix, 63122 Saint Genès Champanelle

²UMR INRA-Oniris 1300 Biologie, Epidémiologie, Analyse de Risque en santé animale
Atlantique, La Chantrerie, B.P. 40706, F-44307 Nantes

1 Principes des différentes méthodes d'échantillonnage des tiques

1.1 Capture des tiques sur la végétation

1.1.1 Méthode du drapeau (MacLeod, 1932, Vassallo et al., 2000)

L'échantillonnage par la méthode du drapeau (« dragging method » en anglais) permet de prélever les tiques à l'affût en simulant le passage d'un hôte. Ce leurre mécanique est réalisé à partir d'une pièce de tissu, appelée drapeau, attachée à un bâton. Ce drapeau est traîné sur la végétation par le préleveur à l'aide d'une corde fixée aux deux extrémités du bâton (Figure 1). Le tissu peut être en éponge, flanelle ou molleton et mesure en général un mètre carré (1 m x 1 m). Pour visualiser plus facilement les tiques, ce tissu est de couleur claire. Le préleveur traîne le drapeau derrière lui, lentement (à environ 0,5 m/s), classiquement sur une distance de 10 m. Une variante consiste à balayer la végétation latéralement (on parle alors de « flagging ») : le drapeau, plus petit, est maintenu par un manche en bois. Au passage du drapeau, les tiques à l'affût s'accrochent et sont ensuite comptées par le préleveur.



Figure 1 : Méthode du drapeau (« dragging ») (Source : S. Bord)

1.1.2 En marchant : tissu qui enserre les jambes du préleveur

Comme pour la méthode du drapeau, l'échantillonnage en marchant (« walking » en anglais) permet de prélever les tiques à l'affût : le leurre en tissu est cette fois-ci placé autour des jambes du préleveur. Cette méthode permet de collecter les tiques sur toute la hauteur de la végétation.

1.1.3 Pièges à CO₂ (Gray, 1985)

L'échantillonnage par piège à CO₂ consiste à attirer les tiques par l'émission de CO₂, qui simule la présence d'un hôte à proximité. Le CO₂ agit comme un agent attractif qui engendre l'activation immédiate des tiques. Ce leurre chimique est réalisé à partir d'une boîte contenant de la neige carbonique déposée sur un support au sol. La boîte est percée pour diffuser le CO₂ et un ruban adhésif double face est collé au support. Les tiques attirées par le dégagement de CO₂ se dirigent vers la boîte et viennent se coller sur le ruban adhésif. Les tiques capturées sont ensuite comptées.

1.1.4 Comptage « à vue » (Pérez-Eid, 2007)

L'échantillonnage à vue consiste à repérer visuellement les tiques à l'affût sur la végétation (en particulier sur les tiges de graminées).

1.1.5 Echantillonnage d'humus (Schulze et al., 1997)

Un prélèvement d'humus permet de récupérer l'ensemble des tiques libres, en recherche d'hôte ou en cours de réhydratation. Les tiques présentes sont séparées de l'humus par flottation, tamisage, ou par la méthode de l'entonnoir « Tullgren », qui consiste à déposer l'humus dans un entonnoir et à placer une lumière allumée au-dessus. La chaleur dégagée par la lumière va sécher la matière et pousser les tiques à se diriger vers le fond du dispositif, sur lequel est déposé un buvard humide. Les tiques seront ensuite comptées sur le buvard.

1.2 Capture des tiques sur hôtes

L'échantillonnage des tiques à partir des hôtes consiste à compter le nombre de tiques présentes sur l'animal. Les animaux sauvages observés sont capturés lors de sessions de piégeage (micromammifères, chevreuils...) ou tués lors de campagnes de chasse (sangliers, chevreuils...). Les animaux d'élevage ou domestiques (bovins, chiens...) peuvent également faire l'objet d'inspections. Cet échantillonnage des tiques peut être réalisé directement sur le terrain, lorsque les animaux doivent être relâchés. Il peut aussi être pratiqué en laboratoire sur des cadavres.

La durée du repas de sang étant relativement courte, l'observation et le comptage des tiques doivent être réalisés rapidement après la capture ou la mort de l'animal pour ne pas sous-estimer l'abondance. Pour réduire ce biais, les animaux sacrifiés sont placés dans des sacs pour que les tiques qui se détacheraient puissent être récupérées et comptées. Généralement, le comptage se limite à quelques parties du corps, considérées comme les zones de fixation privilégiées. Par exemple, l'examen concernera le bec pour les oiseaux, la tête pour les micromammifères, l'aine ou la mamelle pour les bovins (L'Hostis, 1994), ou encore les pattes pour les chevreuils (Gilot et al., 1994).

2 Avantages, contraintes et limites de chaque méthode des différentes méthodes

2.1 Capture des tiques sur la végétation

2.1.1 Méthode du drapeau (MacLeod, 1932, Vassallo et al., 2000)

La méthode du drapeau présente l'avantage d'être facile et rapide à mettre en œuvre, car elle ne nécessite que peu de matériel. Elle est particulièrement adaptée aux espèces de tiques dont la recherche de l'hôte se fait par une quête à l'affût. Chez ces espèces, la méthode est plus

efficace pour les stades dont le site de quête est situé sur le haut de la végétation. Cette méthode est particulièrement recommandée pour les végétations relativement uniformes telles que les prairies (Sonenshine, 1994).

Pour améliorer l'efficacité de cette méthode, certains auteurs modifient le dispositif en fixant un réseau de tuyau en caoutchouc percé dans lequel du CO₂ est diffusé (Gherman et al., 2012). D'autres mêlent la méthode du drapeau et la collecte de tiques en marchant (« walking-flagging » : (Carroll and Schmidtman, 1992).

Par cette méthode, l'effort d'échantillonnage est facilement quantifiable par la durée de l'échantillonnage et/ou la surface prospectée. Cependant, cette méthode ne donne qu'un indicateur d'abondance (présence/absence de tiques, densité faible, modérée ou forte), à ne pas considérer comme une estimation de la vraie taille de la population présente (Sonenshine, 1994).

Le taux d'échantillonnage de cette méthode du drapeau est estimé à 8% de la population totale de la zone échantillonnée pour *Dermacentor variabilis* (Sonenshine, 1994). D'autres auteurs estiment l'efficacité pour la capture des nymphes à 5,9% pour *Ixodes pacificus* (TalleklintEisen and Lane, 2000) et 6,7% pour *Ixodes scapularis* (Daniels et al., 2000). Ces pourcentages annoncés sont toutefois à considérer comme des moyennes, car la part des tiques réellement à l'affût (seules récupérables par cette méthode) dans la population totale des tiques libres fluctue en fonction des conditions d'échantillonnage (météorologie, saison, type de végétation...).

Cette méthode d'échantillonnage n'est par ailleurs pas efficace lorsque la végétation à échantillonner est mouillée, lors de périodes de pluie ou encore le matin en présence de rosée (Sonenshine, 1994). Son efficacité n'est probablement pas identique selon les types de végétation (Ginsberg and Ewing, 1989) : le contact entre le drapeau et la végétation n'est pas optimal pour une végétation haute, hétérogène ou ligneuse.

2.1.2 En marchant : tissu qui enserre les jambes du préleveur

Par rapport à la méthode du drapeau, cette méthode a l'avantage majeur de représenter réellement le risque auquel un individu est exposé lors d'un contact direct (Ginsberg and Ewing, 1989, Dobson et al., 2011). En marchant, l'effort d'échantillonnage est facilement estimable par la durée d'échantillonnage et/ou la distance parcourue par le préleveur (Ginsberg and Ewing, 1989). La méthode permet en outre de mesurer la hauteur d'affût des différentes espèces et des différents stades collectés, dès lors que l'on s'astreint à rechercher fréquemment la présence de tiques sur le tissu, de sorte qu'elles n'aient pas le temps de se déplacer.

L'efficacité de cette méthode peut cependant être biaisée par de nombreux facteurs, comme par exemple un effet préleveur (qui peut attirer plus ou moins les tiques), ou une variabilité de contact avec la végétation (Vassallo et al., 2000).

Selon le même principe, pour estimer le risque lié à un contact direct, certains auteurs utilisent des animaux « sentinelle » (Leschnik et al., 2012).

2.1.3 Pièges à CO₂ (Gray, 1985)

Les pièges à CO₂ permettent de réduire l'effort d'échantillonnage grâce à l'utilisation du ruban adhésif placé autour de la boîte à CO₂. Ces dispositifs basés sur l'attraction sont particulièrement efficaces pour les tiques dont la quête de l'hôte se fait par la chasse, telles que *Hyalomma marginatum*.

L'efficacité des pièges à CO₂ a été démontrée pour les espèces de tiques comme *Amblyomma* spp. (Solberg, 1992, Schulze et al., 1997). Elle diffère selon les espèces qui peuvent être très peu, voire pas du tout collectées par cette méthode (telles que *Dermacentor variabilis*). A l'inverse, d'autres espèces qui recherchent un hôte à l'affût, comme *Ixodes dammini* ou *Ixodes ricinus*, sont facilement collectées par ce type de dispositif (Sonenshine, 1994). Gray (1985) démontré que les pièges à CO₂

sont plus efficaces que la méthode du drapeau pour l'échantillonnage d'*Ixodes ricinus* dans des milieux à haute densité de tiques, et plus particulièrement dans des terrains accidentés et pour les stades adultes.

L'inconvénient majeur des pièges à CO₂ réside dans le fait qu'ils ont un domaine d'efficacité limité (Koch and McNew, 1981, Falco and Fish, 1992). La distance d'attraction a été estimée à 3,5 m pour *Ixodes ricinus* contre plusieurs mètres pour *Amblyomma americanum* (Gray, 1985).

Certaines modifications sont apportées par des auteurs pour améliorer l'efficacité de l'échantillonnage : combinaison du dégagement de CO₂ avec la méthode du drapeau (Gherman et al., 2012), ou encore production de CO₂ par réaction chimique à partir d'acide acétique et de bicarbonate de sodium (Guedes et al., 2012).

2.1.4 Comptage « à vue » (Pérez-Eid, 2007)

La méthode est peu efficace pour les stades larvaires et nymphaux, à cause de leur très petite taille. Le comptage à vue est en revanche plus efficace pour les adultes *Amblyomma* que la méthode du drapeau (Terassini et al., 2010).

2.1.5 Echantillonnage d'humus (Schulze et al., 1997)

L'échantillonnage d'humus permet de collecter l'ensemble des tiques libres présentes sur la zone échantillonnée. Cette méthode ne permet cependant pas de distinguer dans la litière les tiques à l'affût des tiques en cours de réhydratation. De plus, elle est lourde à mettre en œuvre, ce qui la rend difficilement reproductible.

2.2 Capture des tiques sur hôtes

Cette méthode d'échantillonnage peut être réalisée à partir d'animaux de compagnie, domestiques ou sauvages. Dans le cas d'animaux sauvages, ils sont soit tués, soit capturés pour être examinés. Parfois, les animaux vivants nécessitent une anesthésie pour l'examen. Classiquement, il s'agit d'animaux de petite ou de moyenne taille, plus faciles à manipuler. Pour limiter les biais de détection, plusieurs méthodes sont possibles. L'examen peut se faire sur une durée déterminée et fixe, sur une zone privilégiée comme la tête, l'encolure ou les oreilles. L'examen peut également être exhaustif et concerner uniquement une zone circonscrite dont la taille est identique entre les individus. Une autre méthode consiste à placer les animaux capturés dans une cage grillagée au-dessus d'un bac d'eau et attendre que les tiques se détachent de leur hôte naturellement (Mather and Spielman, 1986, Gage et al., 1990, Sonenshine, 1994).

3 Evaluation de l'efficacité de chaque méthode

L'efficacité de chacune de ces méthodes n'a été que très peu quantifiée, voire jamais étudiée. Cependant, certaines études ont été menées pour comparer le nombre de captures obtenues en réalisant des échantillonnages simultanés à partir de plusieurs méthodes.

Il a été démontré que l'efficacité d'une méthode est spécifique à l'espèce et au stade d'intérêt. Cette efficacité est influencée par :

- le comportement des tiques : la hauteur de quête, la stratégie de quête de l'hôte (chasseur actif ou à l'affût), la capacité à rester accroché au drapeau, l'attractivité du CO₂...
- le type d'habitat : capacité du drapeau à entrer en contact avec la végétation...

- les limites des méthodes d'échantillonnage elles-mêmes : le drapeau ne permet pas de détecter les tiques à l'affût sur les strates proches du sol lorsque la végétation est haute, au contraire du walking ; le périmètre d'action est limité pour les pièges à CO₂...

- les conditions climatiques : le vent et l'ensoleillement entraînent une dessiccation des tiques à l'affût, ainsi qu'une mauvaise diffusion du CO₂ ; l'humidité gêne les prélèvements par le drapeau et en marchant.

Prendre en compte les biais de chaque méthode d'échantillonnage est primordial, car ils peuvent influencer l'interprétation des résultats et donc conduire à des connaissances erronées sur l'écologie des tiques (Solberg, 1992, Petry et al., 2010).

4 Références

- Carroll, J.F., Schmidtman, E.T., 1992. Tick sweep - modification of the tick drag-flag method for sampling nymphs of the deer tick (Acari, Ixodidae). *Journal of Medical Entomology* 29, 352-355.
- Daniels, T.J., Falco, R.C., Fish, D., 2000. Estimating population size and drag sampling efficiency for the blacklegged tick (Acari : Ixodidae). *Journal of Medical Entomology* 37, 357-363.
- Dobson, A.D.M., Finnie, T.J.R., Randolph, S.E., 2011. A modified matrix model to describe the seasonal population ecology of the European tick *Ixodes ricinus*. *Journal of Applied Ecology* 48, 1017-1028.
- Falco, R.C., Fish, D., 1992. A comparison of methods for sampling the deer tick, *Ixodes dammini*, in a Lyme disease endemic area. *Exp. Appl. Acarol.* 14, 165-173.
- Gage, K.L., Burgdorfer, W., Hopla, C.E., 1990. Hispid Cotton Rats (*Sigmodon hispidus*) as a Source for Infecting Immature *Dermacentor variabilis* (Acari: Ixodidae) with *Rickettsia rickettsii*. *Journal of Medical Entomology* 27, 615-619.
- Gherman, C., Mihalca, A., Dumitrache, M., Gyorke, A., Oroian, I., Sandor, M., Cozma, V., 2012. CO2 flagging - an improved method for the collection of questing ticks. *Parasites & Vectors* 5, 125.

- Gilot, B., Bonnefille, M., Degeilh, B., Beaucournu, J.-C., Pichot, J., Guiguen, C., 1994. La colonisation des massifs forestiers par *Ixodes ricinus* (Linné, 1758) en France: Utilisation du chevreuil, *Capreolus capreolus* (L. 1758) comme marqueur biologique. *Parasite* 1, 81-86.
- Ginsberg, H.S., Ewing, C.P., 1989. Comparison of flagging, walking, trapping, and collecting from hosts as sampling methods for northern deer ticks, *Ixodes dammini*, and lone-star ticks, *Amblyomma americanum*; (Acari: Ixodidae). *Experimental and Applied Acarology* 7, 313-322.
- Gray, J.S., 1985. A Carbon Dioxide Trap for Prolonged Sampling of *Ixodes ricinus* Populations. *Experimental and Applied Acarology* 1, 35-44.
- Guedes, E., de Azevedo Prata, M., Reis, É., Cançado, P., Leite, R., 2012. Comparative efficiency of two models of CO2 traps in the collection of free-living stages of ixodides. *Parasitology Research* 111, 2325-2328.
- Koch, H.G., McNew, R.W., 1981. Comparative Catches of Field Populations of Lone Star Ticks by CO2-Emitting Dry-Ice, Dry-Chemical, and Animal-Baited Devices. *Annals of the Entomological Society of America* 74, 498-500.
- L'Hostis, M. 1994. *Babesia divergens* en France: Epidémiologie descriptive et analytique. Thèse en parasitologie. Université de Montpellier I, Montpellier.
- Leschnik, M.W., Khanakah, G., Duscher, G., Wille-Piazzai, W., HÖRweg, C., Joachim, A., Stanek, G., 2012. Species, developmental stage and infection with microbial pathogens of engorged ticks removed from dogs and questing ticks. *Medical and Veterinary Entomology* 26, 440-446.
- MacLeod, J., 1932. The bionomics of *ixodes ricinus* L, the "sheep tick" of Scotland. *Parasitology, Cambridge* Vol. 24, pp. 382-400.
- Mather, T.N., Spielman, A., 1986. Diurnal Detachment of Immature Deer Ticks (*Ixodes dammini*) from Nocturnal Hosts. *The American Journal of Tropical Medicine and Hygiene* 35, 182-186.
- Pérez-Eid, C., 2007. Les tiques. Identification, biologie, importance médicale et vétérinaire. Lavoisier, Paris.

- Petry, W., Foré, S., Fielden, L., Kim, H.-J., 2010. A quantitative comparison of two sample methods for collecting *Amblyomma americanum* and *Dermacentor variabilis* (Acari: Ixodidae) in Missouri. *Experimental and Applied Acarology* 52, 427-438.
- Schulze, T., Jordan, R., Hung, R., 1997. Biases associated with several sampling methods used to estimate abundance of *Ixodes scapularis* and *Amblyomma americanum* (Acari: Ixodidae). *J Med Entomol* 34, 615 - 623.
- Solberg, 1992. Quantitative evaluation of sampling methods for *Ixodes dammini* and *Amblyomma americanum* (Acari: Ixodidae). *Journal of Medical Entomology* 29, 451-456.
- Sonenshine, D.E., 1994. *Ecological Dynamics of tick-borne zoonoses*. Oxford.
- TalleklintEisen, L., Lane, R.S., 2000. Spatial and temporal variation in the density of *Ixodes pacificus* (Acari : Ixodidae) nymphs. *Environmental Entomology* 29, 272-280.
- Terassini, F.A., Barbieri, F.S., Albuquerque, S., Szabó, M.P.J., Camargo, L.M.A., Labruna, M.B., 2010. Comparison of two methods for collecting free-living ticks in the Amazonian forest. *Ticks and Tick-borne Diseases* 1, 194-196.
- Vassallo, M., Pichon, B., Cabaret, J., Figureau, C.U., Perez-Eid, C., 2000. Methodology for sampling questing nymphs of *Ixodes ricinus* (Acari : Ixodidae), the principal vector of lyme disease in Europe. *Journal of Medical Entomology* 37, 335-339.

Conditions environnementales, taux d'échantillonnage et indicateurs d'abondance

L'objectif de cette partie est de (i) présenter une approche qui permet de déterminer l'indicateur d'abondance le moins soumis aux variations du taux d'échantillonnage ; (ii) comparer cet indicateur d'abondance à l'indicateur d'abondance usuel *i.e.* basé sur le nombre de tiques capturées lors d'un seul échantillonnage par unité d'observation. Cette approche fait l'objet d'un article en préparation.

Deux campagnes de terrain ont été réalisées en septembre 2011 et mai 2012 au cours desquelles 60 et 58 unités d'observations ont été observées. 12 échantillonnages successifs ont été réalisés sur chacune des unités observations. L'intervalle de temps entre deux échantillonnages successifs était de 2 minutes et 30 secondes.

Nous avons montré par une approche GLM (Generalized Linear Models, McCullagh and Nelder (1989)) que la probabilité pour une tique d'être capturée par le drap au premier échantillonnage était influencée par le type de végétation ($\tau_{\text{Feuilles Mortes}} > \tau_{\text{Arbustes}}$) et l'heure d'échantillonnage ($\tau_{\text{matin}} > \tau_{\text{après-midi}}$). Nous avons également montré qu'en augmentant le nombre d'échantillonnage successifs jusqu'à 4, la probabilité pour une tique d'être capturée était seulement influencée par l'heure d'échantillonnage. Nous en avons donc déduit que l'indicateur d'abondance le moins soumis aux variations du taux d'échantillonnage était l'indicateur basé sur le nombre de tiques capturées lors 4 échantillonnages successifs (AI_4).

Nous avons ensuite cherché à mettre en évidence les facteurs significativement associés à l'indicateur d'abondance usuel (AI_1), puis à l'indicateur d'abondance le moins soumis aux variations du taux d'échantillonnage AI_4 et enfin à l'indicateur le plus proche de la vraie taille de population *i.e.* AI_{12} . Nous avons montré que la saison et l'heure d'échantillonnage étaient significativement associés à AI_4 et AI_{12} ($AI_{\text{matin}} < AI_{\text{après-midi}}$ et $AI_{\text{automne}} > AI_{\text{printemps}}$) alors que seule la saison était associée à AI_1 ($AI_{\text{automne}} > AI_{\text{printemps}}$).

Ces résultats mettent en évidence le fait que l'indicateur d'abondance basé sur un seul échantillonnage n'est pas satisfaisant dans la mesure où il est influencé par les variations du taux d'échantillonnage et qu'il manque de puissance pour mettre en évidence l'effet de facteurs associés à l'abondance. L'avantage de cette approche

est de prendre en compte les facteurs qui influencent le taux d'échantillonnage sans avoir à l'estimer par une modélisation sophistiquée. Cependant, elle ne permet pas de prendre en compte tous les facteurs de variation du taux d'échantillonnage. Cette limite souligne donc l'importance d'estimer simultanément la taille de la population et le taux d'échantillonnage.

Article en préparation

1

1 Consecutive sampling to take into account variation in sampling rate to improve abundance
2 indicator of *Ixodes ricinus*

3

4 S.Bord^{*1}, P.Gasqui¹, P.Druilhet², G.Vourc'h¹

5 ¹ INRA, UR346 Unité Epidémiologie Animale, F-63122 Saint Genès Champanelle, France

6 ² Laboratoire de mathématiques, UMR CNRS 6620 Campus des Cézeaux, B.P. 80026, F-63171
7 Aubière, France

8

9 *Corresponding author: Tel. : +33 4 73 62 42 28; INRA, UR346 Unité Epidémiologie Animale,
10 F-63122 Saint Genès Champanelle, France; e-mail : sbord@clermont.inra.fr

11

12

13

14

15

16

17

Article en préparation

2

18 ABSTRACT

19 The abundance of questing *Ixodes ricinus* ticks in a given observation unit is usually estimated
20 based on the number of nymphs collected during a single drag when using the drag sampling
21 method. This estimate is based on the implicit assumption that nymph capture probabilities are
22 the same across different observation unit, which are characterized by their own sampling
23 conditions. However, this assumption has been questioned by some researchers.

24 The aim of the present study was to determine how to propose one abundance index that
25 minimize variations of sampling rate due to the sampling conditions were taken into
26 consideration. To do so, we dragged for ticks during two different time periods using a removal
27 sampling approach. We investigated the association between sampling conditions, nymph
28 abundance estimates, and the probability that nymphs would be captured during consecutive
29 drags.

30 The probability of capturing a nymph during the first drag was significantly influenced by the
31 vegetation type and the time of day. It was 1.2 times higher in dead leaves than in shrubs and 1.3
32 times higher in the morning than in the afternoon. These results indicate that the abundance
33 estimates typically reported, which are based on the number of captures obtained during a single
34 drag, fail to satisfactorily measure abundance since they vary not only because abundance varies
35 but also because capture probabilities vary with sampling conditions. We show that, in our data,
36 using at least four consecutive drags eliminates vegetation-type-based bias in capture
37 probabilities.

38 **KEYWORDS:** drag sampling method; probability of capture; *Ixodes ricinus*; removal sampling.

39 INTRODUCTION

40 One of the most prevalent vector-borne diseases in the temperate zones of the Northern
41 Hemisphere is Lyme disease (Bacon et al., 2008). In order to estimate the risk of Lyme
42 transmission to humans, it is necessary to estimate the abundance of the tick vector: *Ixodes*
43 *ricinus* (for the most part) in Europe, as well as *I. scapularis* and *I. pacificus* in the USA (Diuk-
44 Wasser et al., 2012).

45 The abundance of *I. ricinus* is usually estimated based on the number of nymphs collected using
46 the drag sampling method (Vassallo et al., 2000). In the drag sampling method, the vegetation
47 within a defined area, called the sub-transect (which generally spans 10 m²) is dragged once
48 (MacLeod, 1932). Host-seeking nymphs, *i.e.* those waiting for a host on top of the vegetation, are
49 collected by the drag. The total number of nymphs collected across several sub-transects is
50 determined to identify associated risk factors. However, this analysis is only valid if the sampling
51 rate, *i.e.* the probability that a tick will be captured (hereafter referred to as “capture probability”)
52 by a single drag, is similar across different types of sub-transects. This major assumption is often
53 discussed in literature (Ginsberg and Ewing, 1989, Solberg, 1992, Petry et al., 2010, Dobson et
54 al., 2011). Using drag sampling, Bord and colleagues (2014) estimated the sampling rate of *I.*
55 *ricinus* nymphs under several types of sampling conditions. They found that sampling rate could
56 be influenced by both the time of day and vegetation type. In this context, the major
57 disadvantages of using a single drag to estimate abundance are the following: a) biased
58 abundance estimates are obtained and b) significant differences in abundance estimates may be
59 mistakenly attributed to other factors when they are actually due to significant differences in
60 capture probabilities (MacKenzie and Kendall, 2002).

Article en préparation

4

61 To address this problem, sampling conditions that influence capture probabilities should be
62 identified. The two main approaches used to study capture probabilities are Capture-Mark-
63 Recapture (CMR) and removal sampling (RS). In CMR, capture probabilities are estimated using
64 the ratio of the number of marked to unmarked individuals. In the case of tick nymphs, this
65 sampling approach would be labor intensive and marked nymphs may demonstrate modified
66 behavior. Removal sampling consists of capturing individuals several times. This approach
67 requires that sampling be conducted over a short period of time to ensure that sampling
68 conditions, such as the weather, remain the same. Removal sampling is easy to use when
69 sampling ticks.

70 The objective of this study was to propose a means of estimating tick abundance for each
71 observation unit that minimizes the influence of sampling conditions on capture probabilities. To
72 do so, we used a removal sampling approach and dragged the environment twelve times in a row.
73 Using the total number of ticks captured during each drag, twelve abundance indices (AIs) were
74 calculated. We investigated the association between sampling conditions and both capture
75 probabilities as well as abundance indices. We discuss the consequences of the associations we
76 found and draw conclusions about the problems involved in estimating tick abundance based on a
77 single drag, as is usually done. We also make recommendations regarding the number of
78 consecutive drags that are necessary to have confidence in tick abundance estimates.

79

80

81

82 MATERIALS AND METHODS

83 *Sampling design and study area*

84 The study was conducted in the Forêt de Sénart, a 3200-ha suburban forest located 22 km
 85 southeast of Paris, France (2°29'36"E, 48°39'24"N). Removal sampling of ticks was conducted
 86 over two time periods or "seasons": September 12–23, 2011, and May 7–18, 2012. We sampled
 87 60 and 58 observation units in 2011 and 2012, respectively. One observation unit was defined as
 88 a delimited area of 1m x 10m and their boundaries were marked. The vegetation in observation
 89 units consisted of either dead leaves or young oaks (*Quercus* spp.) and/or hornbeam (*Carpinus*
 90 *betulus*), which is considered to be a shrub. Data were collected between 8 a.m. and 11 a.m.
 91 ("morning") or between 12 p.m. and 5 p.m. ("afternoon"). Twelve drags were conducted on each
 92 observation unit to collect host-seeking nymphs. The time interval between drags was limited to
 93 two minutes and thirty seconds; it was kept as short as possible to ensure that the population of
 94 host-seeking nymphs remained "closed" over the twelve consecutive drags (the total sampling
 95 duration was less than 30 minutes). Furthermore, if the interval had been too long, environmental
 96 conditions such as weather, temperature, and humidity could have changed. As a result, host-
 97 seeking ticks could have responded differently to the mechanical stimuli created by the dragging.
 98 If they had, sampling rate would no longer have been constant across drags.

99 Nymphs were sampled by dragging a white cloth measuring 1 m x 1 m across surface of the
 100 vegetation located within each 10-m² sub-transect. Nymphs were removed at the end of each drag
 101 and counted. The vegetation type (dead leaves or shrubs) and the time of day (morning or
 102 afternoon) were also recorded. Temperature and relative humidity were measured during each
 103 drag, and the mean values for both were determined for each sub-transect.

Article en préparation

6

104 *Capture Probabilities*

105 To study the influence of sampling conditions on capture probabilities over consecutive drags, we
 106 modeled the probability that a nymph would be captured during the first k drags. For a given
 107 observation unit, capture probabilities were defined as the total number of nymphs captured
 108 during the first k drags divided by the total number of nymphs captured across all drags. Only the
 109 capture probabilities during the first k drags where k varied from one to the total number of drags
 110 and where the capture probability reached 0.8 were modeled.

111 We analyzed the data using GLM quasi-binomial models (McCullagh and Nelder, 1989) with a
 112 logit link (MASS package; R Development Core Team, 2008). In each model, the effects of
 113 season (autumn/spring), vegetation type (dead leaves/shrubs), and time of day
 114 (morning/afternoon) on capture probabilities were tested. Temperature and humidity were not
 115 included in the models since they were both highly correlated with the time of day. Using the
 116 results, we identified the minimum number of drags (k_{MNS}) necessary to eliminate the influence
 117 of sampling conditions on capture probabilities.

118 *Abundance index*

119 Twelve abundance indices, defined as AI_j where $j \in [1;12]$ were calculated by summing the
 120 cumulative number of nymphs captured during the first j consecutive drags. These AI_j s were
 121 considered to be estimates of the size of the population of host-seeking nymphs within a given
 122 observation unit. Usually, in other studies, AI is based on the number of nymphs captured during
 123 a single drag (*i.e.*, AI_1). We considered AI_{122} to be a better estimate of AI since we assumed that
 124 it was a better approximation of the true abundance of host-seeking nymphs. We also explored
 125 whether AI at k_{MNS} was significantly associated with the same sampling conditions as AI_{122} .

Article en préparation

7

126 Furthermore, we compared this result with the standard estimate of AI used in other studies, *i.e.*,
 127 AI_1 .

128 We used negative binomial GLM models using a log link (McCullagh and Nelder, 1989) to
 129 analyze the abundance of host-seeking nymphs as represented by the twelve abundance indices
 130 (AI_j , AI_j where $j \in [1;12]$) (package MASS; R Development Core Team, 2008). In each model, the
 131 effects of season (autumn/spring), vegetation type (dead leaves/shrubs), and time of day
 132 (morning/afternoon) were tested. Temperature and humidity were not included in the models
 133 since they were both highly correlated with the time of day.

134

135 RESULTS

136 An average of 39.7 (SD=27.9) nymphs were captured across the twelve consecutive drags
 137 conducted at each observation unit (10 square meters) ; of all the nymphs caught, 47.9% (SD=
 138 13.8) were captured during the first drag, 61.9% (SD=12.5) had been captured after the first two
 139 drags ($k=2$), 70.3% (SD=10.2) after the first three drags ($k=3$), 77.1% (SD=9.6) after the first
 140 four drags, and 81.6% (SD=9.0) after the first five drags.

141 Patterns of tick exhaustion were not homogeneous. Indeed, three different patterns were
 142 observed: (i) a decline in the number of captures over consecutive drags; (ii) a decline over the
 143 first drags, then an increase in the number of captures, and finally another decline; and (iii) a
 144 slightly declining but relatively constant number of captures (see the figures in the
 145 Supplementary Material).

146

Article en préparation

8

147 *Capture probabilities* (Table 1)

148 Sampling conditions influenced capture probabilities for up to four consecutive drags (k_{MNS}).
 149 Indeed, the probability that a nymph would be captured during the first (s_1), first two (s_2), and
 150 first three (s_3) drags was significantly influenced by both vegetation type and time of day (P -
 151 values < 0.05). The probability that a nymph would be captured was about 1.2 times higher in
 152 dead leaves than in shrubs and 1.25 ($=1/0.8$) times higher in the morning than in the afternoon
 153 (Table 1). Cumulative capture probabilities over the first four (s_4) and five (s_5) drags were only
 154 significantly influenced by the time of day (P -value < 0.05). As a result, for the present data,
 155 conducting at least four drags (s_4) is necessary to control for the influence of vegetation type on
 156 capture probabilities.

157 *Abundance Index* (Table 2)

158 AI_4 (i.e., AI at k_{MNS}) was significantly correlated with season (P -value < 0.001) and time of day
 159 (P -value < 0.05). The same result was observed for the AI_{12} estimate. Estimated nymph
 160 abundance was 1.5 times higher in the autumn than in the spring and 1.3 times higher in the
 161 afternoon than in the morning (Table 2). In the case of AI_{12} , the standard estimate used in most
 162 studies, only season had an influence (P -value < 0.01).

163

164

165 DISCUSSION

166 Being able to accurately estimate questing nymph abundance is of prime importance when it
167 comes to better estimating tick-related health risks. Our results show that conducting several
168 consecutive drags allows us to obtain a far better estimate of tick abundance for each observation
169 unit than just conducting a single drag, which is the approach most commonly used during drag
170 sampling. This is because nymph capture probabilities are minimally influenced by sampling
171 conditions after several drags. In our context, four consecutive drags minimizes the influence of
172 sampling conditions. Furthermore, our results show that the abundance index calculated on the
173 first four drags (AI_4) is as good as the best abundance index available (*i.e.*, the abundance index
174 associated with the full set of twelve drags). Indeed, AI_4 was influenced in the same way by
175 sampling conditions as the best one.

176 Although capture probabilities were no longer biased by vegetation type after four drags, they
177 were still influenced by the time of day. Vegetation type influences capture probabilities because
178 vegetation modifies the contact between the cloth used for dragging and the questing ticks
179 (Milne, 1943, Ginsberg and Ewing, 1989, Sonenshine, 1994, Schulze et al., 1997, Tack et al.,
180 2011). We suspect that increasing the number of drags increases the probability of contact in
181 complex heterogeneous vegetation, like shrubs, such that it reaches the level observed in
182 homogeneous vegetation such as dead leaves. Increasing the number of consecutive drags does
183 not eliminate the effect of time of day on capture probabilities. The time of day could influence
184 the ability of the drag cloth to be in contact with the questing population. The reason is that
185 nymphs would quest at lower levels in the vegetation during the middle part of the day, which
186 was mentioned by Randolph and Storey (1999) and that nymph would react better to the drag
187 stimulus in the afternoon. Our results suggest that nymph performance may be better in the

Article en préparation

10

188 morning than in the afternoon; this pattern may result because night-time temperatures and
189 humidity levels allow ticks to maintain their water balance (Sonenshine, 1994) and thus nymphs
190 may be in better condition in the morning or because, as mentioned by (Randolph and Storey,
191 1999), they stayed high on the vegetation without actively questing. Increasing the number of
192 drags does not seem to erase the difference in capture probabilities of nymphs between the
193 morning and afternoon.

194

195 The abundance index after four drags (AI_4) showed the same relationship with season and time of
196 day as did the abundance index after twelve drags (AI_{12}). The fact that AI_4 was as informative
197 as AI_{12} may be explained by the fact that, on average, 77.1% (SD=9.6) of the nymphs were
198 captured after four drags. Consequently, AI_4 could be considered in our study to be a useful AI. In
199 contrast, the index most commonly used in tick abundance studies (AI_1) is only influenced by
200 season and thus lacks the power to detect the subtle impact of sampling conditions.

201

202 Season is indeed a major factor that explains tick abundance. Ticks are usually more abundant in
203 the spring than in the autumn (Lees and Milne, 1951, Randolph et al., 2002, Perret et al., 2004).
204 Surprisingly, in our study, tick abundance in autumn 2011 was 1.5 fold that in spring 2012. We
205 suspect that, since the period of January–March 2012 was much colder than average (source:
206 Météo France <http://www.meteofrance.com>), the survival of *I. ricinus* sub-adults was diminished
207 (Sonenshine, 1994), and thus spring abundance was also lower. In the case of AI_4 and AI_{12} ,
208 nymphs were 1.3 times more abundant in the afternoon than in the morning. This result has been
209 observed many times (e.g., (Sonenshine, 1994, Greenfield, 2011)) and is generally explained by

Article en préparation

11

210 the fact that temperature, solar radiation, and humidity are optimal for questing nymphs in the
211 afternoon (Sonenshine, 1994) .

212

213 Our study shows that interpretations of abundance estimates for a given observation unit must
214 take into account the effect of sampling conditions on capture probabilities to allow comparisons
215 between several observations units. Indeed, since capture probabilities were higher in the
216 morning than in the afternoon, we would tend to underestimated afternoon abundance if we only
217 sampled in the morning. The difference between tick abundance in the morning and the afternoon
218 was in fact bigger than we had supposed. Consequently, the epidemiological risk associated with
219 ticks is likely to be largely underestimated.

220

221 It would have been ideal to determine the number of drags at which capture probabilities became
222 stable, *i.e.*, were no longer influenced by sampling conditions. However, in our study, where
223 sampling conditions were relatively homogeneous (only two types of vegetation, two times of
224 day, and two seasons), this ideal was not achieved. We thus suspect that capture probabilities are
225 influenced by numerous sampling conditions in large ecological studies (e.g. Boyard et al., 2007,
226 Diuk-Wasser et al., 2012). The more factors that influence capture probabilities, the more
227 complicated it is to obtain good estimates of abundance (MacKenzie and Kendall, 2002, Quinn et
228 al., 2011, Urban et al., 2012).

229

230 In conclusion, an abundance index based on several consecutive drags provides a better estimate
231 of abundance than the commonly employed abundance index that is based on a single drag. We
232 propose two modifications to the drag sampling approach that will increase the accuracy of tick
233 abundance estimates in future studies. First, we propose to sample few observations units that are

Article en préparation

12

234 representative of the different sampling conditions with as many as possible consecutive
235 samplings. The GLM approach would then be used to determine x the number of consecutive
236 drags from which the probability of capture is the least influenced by sampling conditions,. Then,
237 the first potential modification could be to estimate abundance by systematically conducting x
238 consecutive drags per observation units. This solution would yield an abundance index whose
239 capture probabilities are minimally influenced by sampling conditions. However, it is time
240 consuming and sampling conditions may nonetheless continue to influence capture probabilities.
241 The second potential modification could be to estimate the capture probabilities using the x first
242 consecutive drags of the few observations units observed previously that are representative of the
243 different sampling conditions. These estimated probabilities could then be used to weight the
244 single-drag abundance index using a statistical model of abundance. However, this proposed
245 modification would require the use of a more sophisticated statistical abundance model than is
246 typically used.

Article en préparation

13

247 ACKNOWLEDGMENTS

248 This study was partially funded by the EU grant FP7-261504EDENext and has been catalogued
249 by the EDENext Steering Committee as EDENext 204 (<http://www.edenext.eu>). The contents of
250 this publication are solely the responsibility of the authors and do not necessarily reflect the
251 views of the European Commission. We thank the “Tiques et Maladies Tiques” group of the
252 “Réseau Ecologie des Interactions Durables” for discussing our work with us and providing
253 support. We are thankful to the Office National des Forêts for allowing us to work in the Sénart
254 forest. We thank Jean-Louis Chapuis et Benoît Pisanu for their advice. We give very sincere
255 thanks to Elise Vaumourin and Yoan Delorme for her help with data collection. Thanks to Jessica
256 Pearce-Duvet for proofreading the text’s English.

257 REFERENCES

- 258 BACON, R.M., KUGELER, K.J., MEAD, P.S., CENTERS FOR DISEASE, C., PREVENTION, 2008. SURVEILLANCE FOR LYME
259 DISEASE--UNITED STATES, 1992-2006. MORBIDITY AND MORTALITY WEEKLY REPORT. SURVEILLANCE SUMMARIES
260 (WASHINGTON, D.C. : 2002) 57, 1-9.
- 261 BOYARD, C., BARNOUIN, J., GASQUI, P., VOURCH, G., 2007. LOCAL ENVIRONMENTAL FACTORS CHARACTERIZING *IXODES*
262 *RICINUS* NYMPH ABUNDANCE IN GRAZED PERMANENT PASTURES FOR CATTLE. PARASITOLOGY 134, 987-994.
- 263 DIUK-WASSER, M.A., HOEN, A.G., CISLO, P., BRINKERHOFF, R., HAMER, S.A., ROWLAND, M., CORTINAS, R., VOURC'H,
264 G., MELTON, F., HICKLING, G.J., TSAO, J.I., BUNIKIS, J., BARBOUR, A.G., KITRON, U., PIESMAN, J., FISH, D.,
265 2012. HUMAN RISK OF INFECTION WITH BORRELIA BURGDORFERI, THE LYME DISEASE AGENT, IN EASTERN UNITED
266 STATES. AMERICAN JOURNAL OF TROPICAL MEDICINE AND HYGIENE 86, 320-327.
- 267 DOBSON, A.D.M., FINNIE, T.J.R., RANDOLPH, S.E., 2011. A MODIFIED MATRIX MODEL TO DESCRIBE THE SEASONAL
268 POPULATION ECOLOGY OF THE EUROPEAN TICK *IXODES RICINUS*. JOURNAL OF APPLIED ECOLOGY 48, 1017-1028.
- 269 GINSBERG, H.S., EWING, C.P., 1989. COMPARISON OF FLAGGING, WALKING, TRAPPING, AND COLLECTING FROM HOSTS AS
270 SAMPLING METHODS FOR NORTHERN DEER TICKS, *IXODES DAMMINI*, AND LONE-STAR TICKS, *AMBLIOMMA*
271 *AMERICANUM*; (ACARI: IXODIDAE). EXPERIMENTAL AND APPLIED ACAROLOGY 7, 313-322.
- 272 GREENFIELD, B.P.J., 2011. ENVIRONMENTAL PARAMETERS AFFECTING TICK (*IXODES RICINUS*) DISTRIBUTION DURING THE
273 SUMMER SEASON IN RICHMOND PARK, LONDON. BIOSCIENCE HORIZONS.
- 274 LEES, A.D., MILNE, A., 1951. THE SEASONAL AND DIURNAL ACTIVITIES OF INDIVIDUAL SHEEP TICKS (*IXODES RICINUS* L.)
275 PARASITOLOGY 41, 189-208.
- 276 MACKENZIE, D.I., KENDALL, W.L., 2002. HOW SHOULD DETECTION PROBABILITY BE INCORPORATED INTO ESTIMATES OF
277 RELATIVE ABUNDANCE? ECOLOGY 83, 2387-2393.
- 278 MACLEOD, J., 1932. THE BIONOMICS OF *IXODES RICINUS* L, THE "SHEEP TICK" OF SCOTLAND. PARASITOLOGY, CAMBRIDGE
279 VOL. 24, PP. 382-400.
- 280 MCCULLAGH, P., NELDER, J.A., 1989. GENERALIZED LINEAR MODELS. CHAPMAN & HALL/CRC MONOGRAPHS ON
281 STATISTICS & APPLIED PROBABILITY,
- 282 MILNE, A., 1943. THE COMPARISON OF SHEEP-TICK POPULATIONS (*IXODES RICINUS* L.). ANNALS OF APPLIED BIOLOGY 30,
283 240-250.
- 284 PERRET, J.-L., RAIS, O., GERN, L., 2004. INFLUENCE OF CLIMATE ON THE PROPORTION OF *IXODES RICINUS* NYMPHS AND
285 ADULTS QUESTING IN A TICK POPULATION. JOURNAL OF MEDICAL ENTOMOLOGY 41, 361-365.
- 286 PETRY, W., FORÉ, S., FIELDEN, L., KIM, H.-J., 2010. A QUANTITATIVE COMPARISON OF TWO SAMPLE METHODS FOR
287 COLLECTING *AMBLIOMMA AMERICANUM* AND *DERMACENTOR VARIABILIS* (ACARI: IXODIDAE) IN MISSOURI.
288 EXPERIMENTAL AND APPLIED ACAROLOGY 52, 427-438.
- 289 QUINN, J.E., BRANDLE, J.R., JOHNSON, R.J., TYRE, A.J., 2011. APPLICATION OF DETECTABILITY IN THE USE OF INDICATOR
290 SPECIES: A CASE STUDY WITH BIRDS. ECOLOGICAL INDICATORS 11, 1413-1418.
- 291 RANDOLPH, S.E., GREEN, R.M., HOODLESS, A.N., PEACEY, M.F., 2002. AN EMPIRICAL QUANTITATIVE FRAMEWORK FOR
292 THE SEASONAL POPULATION DYNAMICS OF THE TICK *IXODES RICINUS*. INTERNATIONAL JOURNAL FOR PARASITOLOGY
293 32, 979-989.
- 294 RANDOLPH, S.E., STOREY, K., 1999. IMPACT OF MICROCLIMATE ON IMMATURE TICK-RODENT HOST INTERACTIONS (ACARI :
295 IXODIDAE): IMPLICATIONS FOR PARASITE TRANSMISSION. JOURNAL OF MEDICAL ENTOMOLOGY 36, 741-748.
- 296 SCHULZE, T.L., JORDAN, R.A., HUNG, R.W., 1997. BIASES ASSOCIATED WITH SEVERAL SAMPLING METHODS USED TO
297 ESTIMATE ABUNDANCE OF *IXODES SCAPULARIS* AND *AMBLIOMMA AMERICANUM* (ACARI : IXODIDAE). JOURNAL OF
298 MEDICAL ENTOMOLOGY 34, 615-623.
- 299 SOLBERG, 1992. QUANTITATIVE EVALUATION OF SAMPLING METHODS FOR *IXODES DAMMINI* AND *AMBLIOMMA*
300 *AMERICANUM* (ACARI: IXODIDAE). JOURNAL OF MEDICAL ENTOMOLOGY 29, 451-456.
- 301 SONENSHINE, D.E., 1994. ECOLOGICAL DYNAMICS OF TICK-BORNE ZOONOSES. OXFORD.

Article en préparation

15

- 302 TACK, W., MADDER, M., DE FRENNE, P., VANHELLEMONT, M., GRUWEZ, R., VERHEYEN, K., 2011. THE EFFECTS OF
303 SAMPLING METHOD AND VEGETATION TYPE ON THE ESTIMATED ABUNDANCE OF Ixodes ricinus TICKS IN FORESTS.
304 EXPERIMENTAL AND APPLIED ACAROLOGY 54, 285-292.
- 305 URBAN, N.A., SWIHART, R.K., MALLOY, M.C., DUNNING JR, J.B., 2012. IMPROVING SELECTION OF INDICATOR SPECIES
306 WHEN DETECTION IS IMPERFECT. ECOLOGICAL INDICATORS 15, 188-197.
- 307 VASSALLO, M., PICHON, B., CABARET, J., FIGUREAU, C.U., PEREZ-EID, C., 2000. METHODOLOGY FOR SAMPLING
308 QUESTING NYMPHS OF Ixodes ricinus (ACARI : Ixodidae), THE PRINCIPAL VECTOR OF LYME DISEASE IN EUROPE.
309 JOURNAL OF MEDICAL ENTOMOLOGY 37, 335-339.
- 310
- 311

1 **Tables**

2

3 Table 1. Results of quasi-binomial models examining differences in the capture probabilities of *Ixodes ricinus* nymphs after the first drag (S1), the first
4 two drags (S2), the first three drags (S3), the first four drags (S4), and the first five drags (S5).

5 For each variable, results include the 95% confidence interval (in brackets). The symbols '*', '**', and '***' indicate significance at the P< 0.05, P<
6 0.01, and P<0.001 levels, respectively (e.g., on average, the capture probabilities after the first pass are 1.22 higher in dead leaves than in shrubs)"

7

8

9

	S1	S2	S3	S4	S5
Season					
Spring (reference)	1	1	1	1	1
Autumn	0.96 [0.81-1.14]	1.04 [0.88-1.23]	1.07 [0.91-1.26]	1.08 [0.91-1.28]	1.10 [0.91-1.34]
Vegetation					
Shrubs (reference)	1	1	1	1	1
Dead Leaves	1.22 [1.02-1.44]***	1.22[1.03-1.43]***	1.18 [1.00-1.38]***	1.12 [0.94-1.32]	1.05 [0.87-1.26]
Time of day					
Morning (reference)	1	1	1	1	1
Afternoon	0.82 [0.69-0.98]***	0.74 [0.62-0.88]***	0.77 [0.65-0.91]***	0.81 [0.68-0.97]***	0.80 [0.66-0.98]***

10

11

12

13

14

15 Table 2. Results of the negative binomial models examining differences in *Ixodes ricinus* nymph
 16 abundance. The abundance indices provided here were calculated based on the total number of
 17 nymphs captured after the first drag (AI₁), the first four drags (AI₄), and all twelve drags (AI₁₂),
 18 which we considered to be the best representation of the true abundance of host-seeking nymphs).
 19 For each variable, results include the 95% confidence interval (in brackets). The symbols ‘*’, ‘**’,
 20 and ‘***’ indicate significance at the P< 0.05, P< 0.01, and P<0.001 levels, respectively (e.g.,
 21 nymph abundance after the first drag (AI₁) is 1.46 times higher in the autumn than in the spring.)
 22

23

	AI ₁	AI ₄	AI ₁₂
Season			
Spring (reference)	1	1	1
Autumn	1.46 [1.15-1.85]*	1.51 [1.22-1.87]***	1.48 [1.20-1.83]***
Vegetation			
Shrubs (reference)	1	1	1
Dead Leaves	1.02 [0.803-1.30]	0.97 [0.78-1.20]	0.95 [0.76-1.17]
Time of sampling			
Morning (reference)	1	1	1
Afternoon	1.21 [0.95-1.55]	1.28 [1.03-1.59]*	1.34 [1.08-1.65]**

24

Supplementary material

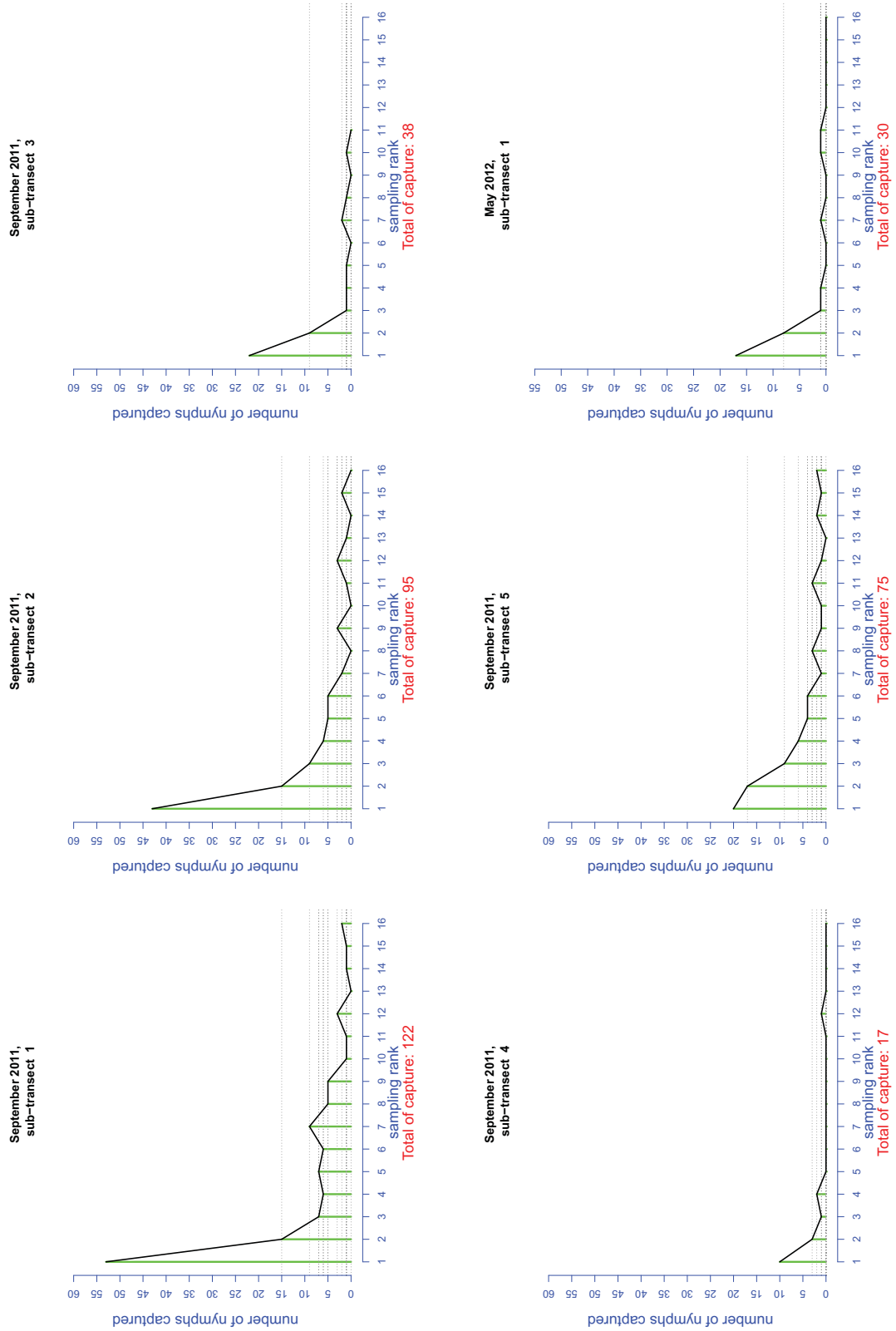


Fig. 3.1 : Examples of decline in the number of captures over consecutive drags

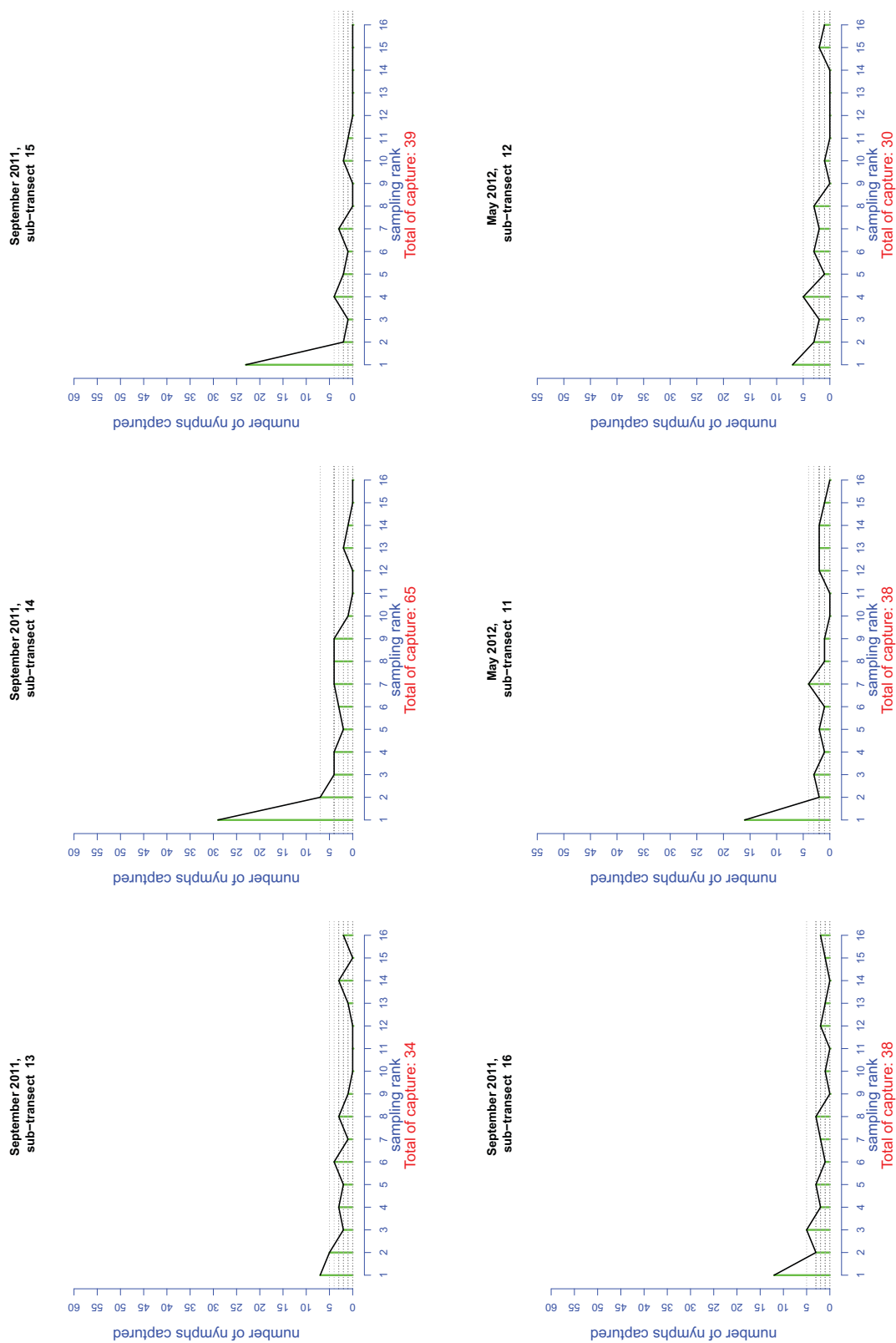


Fig. 3.2 : Examples of aslightly declining but relatively constant number of captures

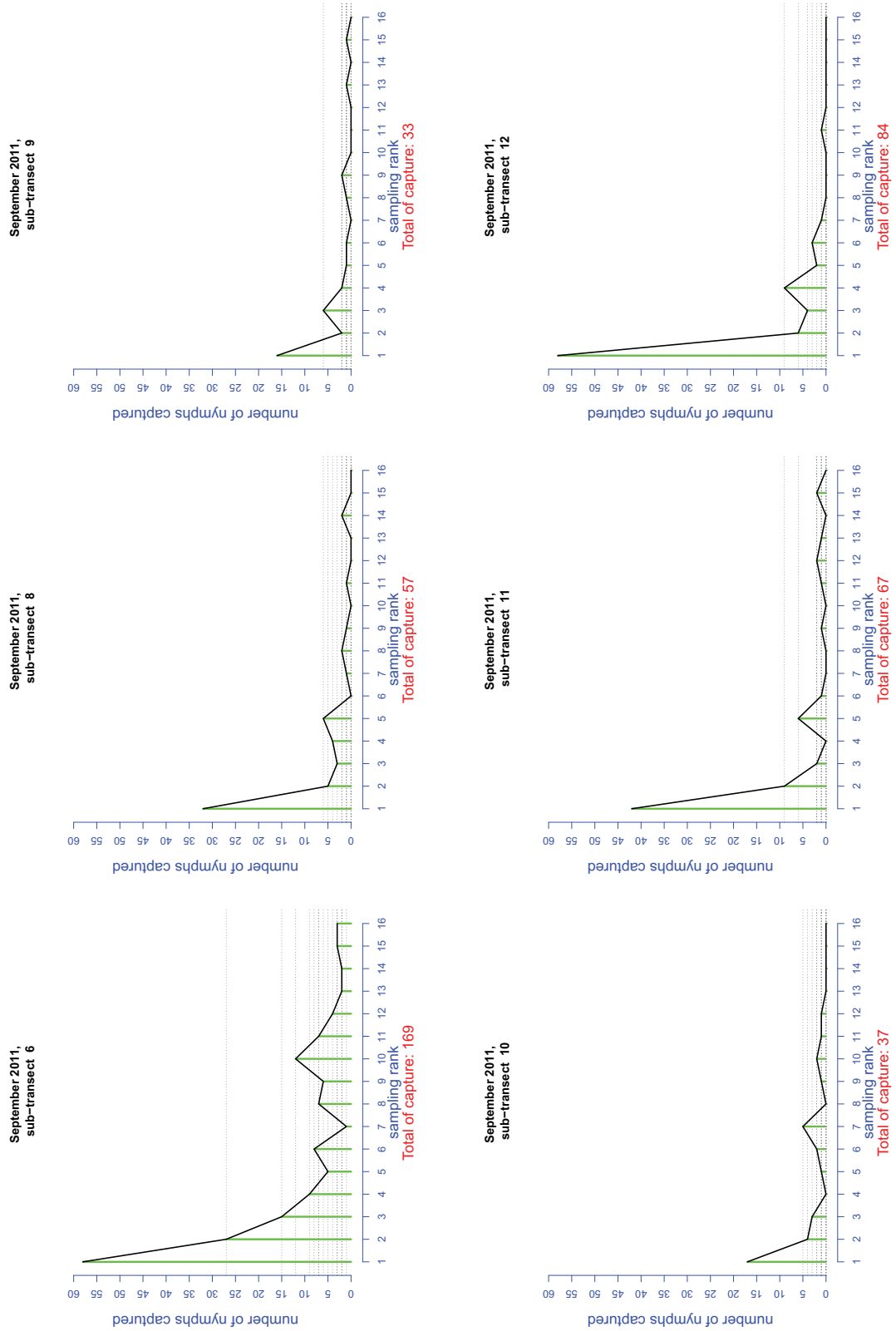


Fig. 3.3 : Examples of decline over the first drags, thus an increase in the number of captures

Modélisation Bayésienne hiérarchique de la taille de la population et du taux d'échantillonnage

L'objectif de cette étude est de proposer une approche bayésienne hiérarchique pour estimer simultanément la taille de la population et le taux d'échantillonnage. Cette partie présente un article publié dans la revue « Ecological Modelling ».

A partir de simulations, nous avons comparé la performance de l'approche bayésienne proposée à celle de l'estimation par la méthode de régression linéaire (Hayne, 1949) qui est utilisée pour estimer le taux d'échantillonnage des tiques par Talleklint-Eisen and Lane (2000).

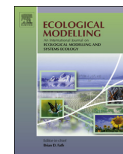
Nous avons montré que l'approche bayésienne estimait mieux la taille de la population que la régression linéaire et que l'estimation du taux était comparable entre les 2 approches. À partir des données récoltées lors de la campagne de septembre 2011, nous avons estimé simultanément la taille de la population et le taux d'échantillonnage en prenant en compte l'effet du type de végétation et l'heure d'échantillonnage en tant qu'effet fixe et nous avons introduit un effet aléatoire pour modéliser une variation aléatoire du taux lié à l'unité d'observation.

Nous avons estimé un taux d'échantillonnage compris entre 33,9% et 47,4% pour les arbustes et compris entre 53,6% et 66,7% pour les feuilles mortes.



Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

Bayesian estimation of abundance based on removal sampling under weak assumption of closed population with catchability depending on environmental conditions. Application to tick abundance

S. Bord^{a,*}, P. Druilhet^b, P. Gasqui^a, D. Abrial^a, G. Vour'h^a^a INRA, UR346 Unité Epidémiologie Animale, Centre INRA de Theix, F-63122 Saint Genès Champanelle, France^b Laboratoire de mathématiques, UMR CNRS 6620 Campus des Cézeaux, B.P. 80026, F-63171 Aubière cedex, France

ARTICLE INFO

Article history:

Received 8 July 2013

Received in revised form

15 November 2013

Accepted 1 December 2013

Available online 28 December 2013

Keywords:

Abundance

Sampling-rate

Removal sampling

Sampling design

Hierarchical Bayesian approach

Hayne method

ABSTRACT

The estimation of animal abundance is essential to understand population dynamics, species interactions and disease patterns in populations. Estimations of relative abundance classically are based on a single observation of several sites. In this case, the mapping of abundance assumes that the probability of detecting an individual, hence the sampling rate, remains constant across the observed sites. In practice, however, this assumption is often not satisfied as the sampling rate may fluctuate between sites due to random fluctuations and/or fluctuations associated with the sampling process, notably associated with the characteristics of the site. It is therefore important to account for variations in detection probability. Using a removal sampling design, we studied the performance of a Bayesian approach to estimate both sampling rates and abundance under the assumption of a closed population. The assumption of a closed population often is weakened when the number of successive samplings is large. The number of samplings has to be limited and optimal. We therefore examined the minimal number of successive samplings needed to achieve sufficient statistical accuracy while respecting underlying model assumptions. Using the same simulations, we also compared the performance of the Bayesian approach to the performance of the frequentist Hayne method based on linear regression. We show that the Bayesian approach proposed gives generally better estimations of population size than the Hayne method. The two methods give approximately the same results for the estimation of sampling rate. We then studied the variability of detection probability of *Ixodes ricinus* ticks sampled under several environmental conditions by using a hierarchical Bayesian model with a random effect. The estimated sampling rate \hat{p}_c varied between 33.9% and 47.4% for shrubs and 53.6% and 66.7% for dead leaves. The variability of the sampling rate due to the site decreased when the number of successive samplings considered in the model increased. The variability was lower in dead leaves than shrubs. This approach could be used routinely for ecological or epidemiological studies of ticks and species with comparable life histories.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The estimation of animal abundance is essential in ecology to understand fundamental processes, such as population dynamics and species interactions, as well as in epidemiology to understand and generate disease patterns in populations (Anderson, 1991). In the majority of biological systems, relevant indicators of abundance are based on count point surveys (Alldredge et al., 2007) obtained using convenient and calibrated sampling methods (Anderson, 2001; Pollock et al., 2002). As a part of the population is often not observable, the probability of detecting an individual and the sampling rate are both less than 1 (Kery, 2004; Kery and

Royle, 2010; Pellet and Schmidt, 2005). Consequently, indicators calculated in this way only give an index of relative abundance. These indicators of abundance are implicitly based on the assumption that the sampling rate is constant from site to site (Williams et al., 2002; Pollock et al., 2002; Royle and Dorazio, 2006). However, the sampling rate may depend on environmental conditions, such as weather, season, sampler and habitats. If this is the case, considering the sampling rate to be constant leads to confusion between the variability of the rate and the variability of abundance (Thompson et al., 1998; MacKenzie and Kendall, 2002). Therefore, an effort needs to be made to estimate both the abundance and the sampling rate. The two main sampling methods used to estimate both abundance and sampling rate are Capture Mark Recapture (CMR) and removal sampling (RS) (see e.g. Williams et al., 2002; Thompson et al., 1998). CMR involves capturing a random sample of individuals in a population of interest, marking them, and then

* Corresponding author. Tel.: +33 4 73 62 42 28.
E-mail address: sbord@clermont.inra.fr (S. Bord).

releasing them back into the population. Another sample of individual from the same population is subsequently recaptured (Williams et al., 2002). The ratio of unmarked to marked individuals is used to estimate capture probability and abundance. CMR is generally labour intensive (Pollock et al., 2002). Moreover, the capture, marking and release steps could have an effect on the behaviour of the marked population (Dodd and Dorazio, 2004), especially in some species. RS consists of capturing individuals on several occasions over a short period of time. The captured individuals are removed from the population (Williams et al., 2002). The total number of individuals captured over the successive samplings is used to estimate capture probability and abundance. Like CMR, RS is labour intensive (Dodd and Dorazio, 2004) and may modify the observed site when the number of successive samplings is high. The choice of a given protocol (CMR or RS) and its ease of implementation depend on the species studied. Although available, CMR and RS methods are rarely used for certain species. One such species are ticks, which are the most important vectors of human and animal diseases after mosquitoes (Parola and Raoult, 2001). The classical index of tick abundance used is estimated by the number of tick nymphs by dragging a piece of cloth once over the vegetation of a delimited area, generally 10 m² (Vassallo et al., 2000) in a selection of sites. Host-seeking nymphs, i.e. those waiting for a host on the top of the vegetation, are collected by the drag. The numbers of nymphs collected on the different sites are then compared. The drag method is distinguished from RS methods in that the cloth is dragged only once over each site. To our knowledge, the sampling rate of the drag sampling method has been studied little. Only one study (Talleklint-Eisen and Lane, 2000) has used a RS design to estimate the abundance and the sampling rate of the drag method. In this study, 17 successive samplings were conducted over 23 days. This protocol could not guarantee that the population remained closed over the sampling period. The authors estimated the sampling rate to be 5.9% using the Hayne method (Hayne, 1949).

To estimate parameters, the Hayne method makes a linear regression of the number of successive captures on the cumulative number of captures. The sampling rate and the population size are estimated respectively as the slope of the regression line and as the intersection point between the horizontal axis and regression line. However, the Hayne method is known to produce poor estimations of population size (White et al., 1982), especially when the sampling rate is potentially low (less than 10%) and variable. Moreover, this method does not allow covariate effects to be taken into account, nor does it provide confidence intervals for estimates.

In this paper, a hierarchical Bayesian approach was used to estimate both the sampling rate and the population size. This approach allows the inclusion of prior knowledge and provides posterior distributions of parameter estimates. Moreover, because it is hierarchical, one can take into account parameters which are either observable or not observable, and which are located at different scales such as the area sampled and the site of sampling (Gelman and Hill, 2006; Cressie et al., 2009). First, we studied the performance of a Bayesian approach using simulations to determine the minimum number of successive samplings needed to achieve sufficient statistical accuracy for estimates while respecting an underlying model assumption of a closed population. We then compared the performance of the Bayesian approach to the Hayne method. Finally, our Bayesian approach was applied to RS tick data collected using the drag method on several sites in September 2011. In a first hierarchical Bayesian model (HBM), the sampling rate was assumed to be specific to each site. In a second HBM, the sampling rate was assumed to be the same for similar conditions with small uncontrolled spatial variations due to the characteristics of the site observed.

2. Materials and methods

2.1. Removal sampling data structure

The statistical unit considered in this study was the sampling site. A sampling site was defined as a delimited area on which successive samplings were performed. For a given sampling site, we considered a closed population of an unknown size N_0 , i.e. with no immigration, emigration, birth or death during the successive samplings. We denoted by $X = (X_1, \dots, X_K)$ the sequence of captures observed on a given site where K was the total number of successive samplings carried out. Let X_k be the number of captures at the k th sampling and N_k the remaining population after the k th sampling where $N_k = N_{k-1} - X_k$ for $k \in 1, \dots, K$. Furthermore, each individual of the closed population was assumed to be captured independently with the same probability of capture τ (Moran, 1951; Zippin, 1956). Hence, we assumed that X_k followed a binomial distribution with population size N_{k-1} and probability of capture τ (Eq. (1)):

$$(X_k | N_{k-1}, \tau) \sim \mathcal{B}(N_{k-1}, \tau), \quad \text{where } N_k = N_{k-1} - X_k. \quad (1)$$

The capture probability was considered as independent and the same for all individuals, so we considered that it was equal to the sampling rate, i.e. the percent of captured individuals in the population.

2.2. Hierarchical Bayesian models

A first HBM (HBM1) assumed that the population size N_{0s} and the sampling rate τ_s was specific to each site s . A second HBM (HBM2) assumed that the sampling rate of a given site τ_s was associated to both the effect of sampling conditions c (considered to be a fixed effect) and a small variation due to the site sampled (considered to be a random effect). The logit transformation of sampling rate τ_s denoted $\text{logit}(\tau_s)$ was used. The $\text{logit}(\tau_s)$ was decomposed as the sum of the logit transformation of the sampling rate τ_c under the sampling conditions c and a random effect ϵ_s (Eq. (2)):

$$\text{logit}(\tau_s) = \text{logit}(\tau_c) + \epsilon_s. \quad (2)$$

The random effect ϵ_s was introduced to add a fluctuation of the sampling rate τ_s due to the observed site s . The range of variation of the random effect ϵ_s , denoted by σ_c^2 , was considered to be specific to each sampling condition c where ϵ_s was assumed to follow a normal distribution with zero-mean and variance σ_c^2 which depended on the sampling conditions c . The HBM1 and HBM2 models described above are summarised in Figs. 1 and 2 by Directed Acyclic Graphs (DAGs) (Thulasiraman and Swamy, 1992; Clark and Gelfand, 2006). These DAGs represent relationships (lines) between the observed data and the unknown parameters or hypotheses of the model (nodes). The lines represent the relations and the hierarchy between nodes. The nodes symbolise the data observed X_k , parameters to estimate N_{0s} , τ_s , τ_c , prior distribution of parameters to estimate for N_{0s} , τ_s , τ_c and ϵ_s and the distribution of hyperparameter σ_c^2 . If $\epsilon_s = 0$ and τ_c was specific to each site s , then the DAG corresponded to the HBM1 model. HBM1 and HBM2 were implemented in OpenBUGS (version 3.2.1.). Source codes are available in the supplementary material.

The prior distribution is a key part of Bayesian inference and represents information about an uncertain parameter. In the absence of accurate information, a uniform prior distribution over the interval $[0; 1000]$ was assigned to N_{0s} . For HBM1, a uniform prior distribution over the interval $[0; 1]$ was assigned to τ_s . For HBM2, a uniform prior distribution over the interval $[0; 1]$ was assigned to τ_c and a uniform prior distribution over the interval $[0; 3]$ was assigned to σ_c^2 . The prior distribution was then combined with the probability distribution of data to obtain the posterior distribution of parameters. Parameters were estimated by the mean and

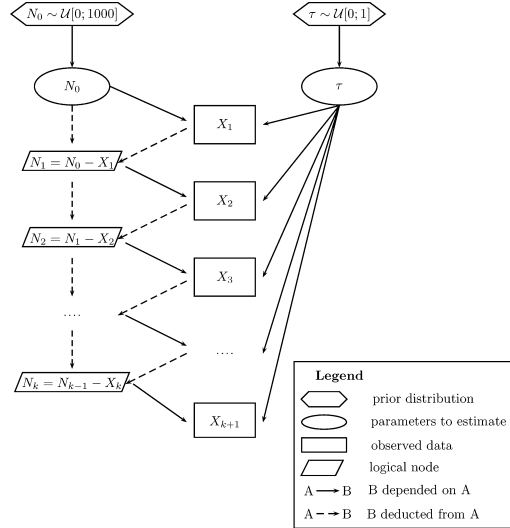


Fig. 1. Directed Acyclic Graph for the hierarchical Bayesian model denoted by HBM1 and *a priori* distributions of parameters N_0 and τ . N_0 represents the size of population; τ represents the sampling rate; X_k represents the observed number of capture at the k th sampling; N_k represents the size of population after the k first samplings.

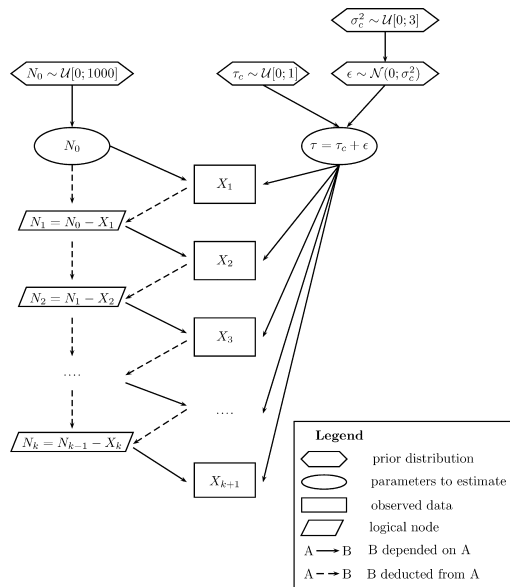


Fig. 2. Directed Acyclic Graph for the hierarchical Bayesian model denoted by HBM2 and *a priori* distributions of parameters N_0 , τ_c , ϵ and σ_ϵ^2 . N_0 represents the size of population; τ represents the sampling rate of a given site which is the sum of the sampling rate associated with the sampling conditions of the given site denoted by τ_c and ϵ which represent the variation of the sampling rate due to the given site; σ_ϵ^2 represents the range of the variation of the sampling rate associated to the conditions ϵ ; N_0 represents the size of population; X_k represents the observed number of capture at the k th sampling; N_k represents the size of population after the k first samplings.

the median of the posterior distribution. To minimize the effect of initial values on the posterior inference, the first 500,000 iterations were discarded as burn-in. The posterior distribution of the parameters was described by thinning three Markov chains and keeping every 50th of the last 500,000 values (Gelman and Shirley, 2011). The convergence of MCMC sampling was visually analysed by inspecting trace and autocorrelation plots.

2.3. Model checking for observed data

To check whether the model fit the data (Gelman et al., 2004), we tested whether the simulations denoted by $X_{\text{rep}} = (X_{\text{rep},1}, \dots, X_{\text{rep},K})$ generated under the model were similar to the observed data denoted by $X = (X_1, \dots, X_K)$. Consequently, we simulated 25,000 sequences of removal sampling X_{rep} from the posterior predictive distribution. The observed data were compared to the simulated X_{rep} through predictive bands of overall coverage level $1 - \alpha$, for some $0 < \alpha < 1$. The shape of the band was constructed by using equal-tailed predictive intervals for each sampling rank $k \in 1, \dots, K$ with the same coverage level $1 - \alpha_{\text{rank}}$. More precisely, for a given α_{rank} , the inferior and superior limits of the Predictive Interval for rank k , denoted by PI_k , were given by the $\alpha_{\text{rank}}/2$, resp. $1 - \alpha_{\text{rank}}/2$, quantile of the 25,000 simulated values $X_{\text{rep},k}$. The overall level $1 - \alpha$ of the predictive band was then estimated by the proportion of X_{rep} inside the band, i.e. the X_{rep} such that each $X_{\text{rep},k}$ belongs to PI_k . The value of α_{rank} was such that the overall risk was equal to α . The validity of the model was accepted if the proportion of the observed sequences $X = (X_1, \dots, X_K)$ that belonged to the predictive band was at least $1 - \alpha$.

2.4. Simulation scheme

To compare the performance of the Hayne method to that of HBM1, simulations were performed. Different levels of population size N_0 and sampling rate τ were tested to evaluate their impact on the quality of estimators. Three population sizes N_0 were considered: 50, 100 and 500. Four values of τ were tested: 0.10, 0.30, 0.50, 0.70. For each combination of N_0 and τ values, we simulated 1000 sequences of removal sampling X_1, \dots, X_{10} according to the model described in HBM1 (Eq. Eq. (1)). The first k successive samplings of each simulated sequence were considered to estimate N_0 and τ , where k ranged from 2 to 10. The combination of N_0 , τ and k values were considered as one scenario. In the end, 108 scenarios were constructed. Based on the 108 scenarios and the 1000 sequences simulated, the estimation of N_0 and τ were performed by both the Hayne method and HBM1.

2.5. Performance of estimators

The classical criterion used to evaluate the performance of an estimator $\hat{\theta}$ of a parameter θ is the Mean Square Error denoted by $MSE_{\hat{\theta}}(\theta)$ (Lehmann and Casella, 1998). $MSE_{\hat{\theta}}(\theta)$ has two components: the variability of the estimator and the distance between the estimator's expected value and the true value of the parameter being estimated or bias (Eq. (3) and (4)):

$$MSE_{\hat{\theta}}(\theta) = \mathbb{E}((\hat{\theta} - \theta)^2) = (\text{bias}(\hat{\theta}))^2 + \text{var}(\hat{\theta}) \quad (3)$$

where

$$\text{bias}(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta \quad (4)$$

The Root Mean Square Error (RMSE), which is often preferred, has the same units of measurement as the quantity being estimated. To have a comparable criterion between the quantities being estimated, the Ratio of the Root of Mean Square Error (RRMSE) to the true value of the parameter estimated was calculated.

For a given value of N_0 , the estimator \hat{N}_0 with the smallest RRMSE is considered to be the best (similarly for τ and its estimator $\hat{\tau}$). An estimator was considered to be performant when the RRMSE was lower than twenty percent (Pollock et al., 1990). We then examined the number of successive samplings that gave a satisfactory quality of estimate. For each scenario (N_0, τ, k) of the simulation study (Section 2.4), the RRMSE values were calculated for N_0 and τ obtained by the Hayne method and the Bayesian approach. They were denoted respectively by $\text{RRMSE}(\hat{N}_0)$ and $\text{RRMSE}(\hat{\tau})$. The performance of the Hayne method and the hierarchical Bayesian approach were compared using the RRMSE values.

2.6. Case study: abundance estimation of *Ixodes ricinus* host-seeking nymphs

Ixodes ricinus host-seeking nymphs were collected from the 15th to the 23th of September 2011 in the "Parc de la Faisanderie" located in the Forêt de Sénart, France. The vegetation on the observed sites was composed of either dead leave or young oak (*Quercus*) and/or hornbeam (*Carpinus betulus*), considered to be "shrub". During the sampling period, 60 sites were sampled either between 8 a.m. and 11 a.m. ("morning") or between 12 p.m. and 5 p.m. ("afternoon"). The 60 sites therefore fell into 4 categories: "shrub sampled in the morning" ($n=8$), "shrub sampled in the afternoon" ($n=15$), "dead leaves sampled in the morning" ($n=17$), "dead leaves sampled in the afternoon" ($n=20$).

Nymphs were captured using the drag sampling method described by Vassallo et al. (2000), which consists of dragging a $1\text{ m} \times 1\text{ m}$ cloth over the vegetation for 10 m. Unfed host-seeking nymphs respond to the mechanical stimuli created by the sweep of the drag cloth by attaching themselves to it. Each 1×10 meter sampling site was marked with boundary markers. For each sampling site, 16 successive removal samplings were realised to collect host-seeking nymphs. One sampling was made every two minutes and thirty seconds. The number of nymphs captured at each sampling was recorded.

The variable of interest was the sequence of consecutive numbers of nymphs captured during the successive samplings on a given observation site. Only the first n samplings were kept for the estimation of the host-seeking population size and the sampling rate. n values varied between 3 and 5 to avoid the emergence of uncontrolled phenomena that could contradict the hypothesis of a closed population such as a modification of environmental conditions or the arrival of new host-seeking ticks (Sonenshine, 1994). Data were modelled through HBM1 and HBM2. The posterior distribution of the parameters was by thinning three Markov chains and keeping every 50th of the last 500,000 values (burn-in was set to 500,000). The sampling rate τ and the size of population N_0 were estimated by the mean of the posterior distribution.

To compare the estimated sampling rates of two different conditions c_1 and c_2 , we monitored the posterior distribution of contrasts. The sampling rate of two conditions c_1 and c_2 are considered to be the same when 0 belongs to the 95% equal tail credible set of contrast $\tau_{c_1} - \tau_{c_2}$.

3. Results

3.1. Performance of estimators

When the number of successive samplings k increased, the $\text{RRMSE}(\hat{N}_0)$ and $\text{RRMSE}(\hat{\tau})$ of the Bayesian approach decreased (see Table 1 and Tables A.1 and A.2 in the supplementary material for details). This result indicates that the estimation is better if the number of successive samplings k considered is larger.

To obtain a satisfactory estimation for $N_0 = 50$, i.e. a RRMSE lower than 0.2, the minimal number of successive samplings should be at least equal to 8, 4 and 3 for $\tau = 0.3, 0.5$ and 0.7 , respectively. For $N_0 = 100$, the minimal number of successive samplings to be considered should be at least equal to 6, 3 and 2 for $\tau = 0.3, 0.5$ and 0.7 respectively. For $N_0 = 500$, the minimal number of successive samplings to be considered should be at least equal to 9, 3 for $\tau = 0.1, 0.3$ respectively and to 2 for $\tau = 0.5$ and 0.7 . At an equal number of successive samplings k , the RRMSE of the Bayesian approach decreased when the true values of N_0 and τ increased. This result indicated that estimators were better for higher values of N_0 and τ .

The $\text{RRMSE}(\hat{N}_0)$ was generally lower for the HBM1 estimator than the Hayne method. The $\text{RRMSE}(\hat{\tau})$ was globally comparable for both methods although it was generally lower for the HBM1 estimator. For $\tau \leq 0.3$, the $\text{RRMSE}(\hat{N}_0)$ and $\text{RRMSE}(\hat{\tau})$ of the HBM1 estimator were globally greater than the Hayne estimator. When the true value of N_0 was equal to 50 or 100, the $\text{RRMSE}(\hat{N}_0)$ of the Hayne estimator was very large.

3.2. Bayesian model estimations for *I. ricinus* abundance

The comparison of the observed data with the predictive bands of an overall coverage level of 95% showed that one sequence $X = (X_1, \dots, X_k)$ over the 60 observed ones was outside the predictive bands for $k=3$. For $k=4$ and $k=5, 6$ and 9 observed sequences were outside the predictive bands. The validity of the model considering the first 3 samplings was thus accepted at a 95% confidence level.

The results presented in the preceding section demonstrate that for a population size of $N_0 = 50$ and a sampling rate $\tau = 0.5$, 4 successive samplings are needed to achieve a good estimation quality. Three models were therefore considered using the first three, four and five successive samplings for the application.

The estimated sampling rate $\hat{\tau}_c$ varied between 33.9% and 47.4% for shrubs and 53.6% and 66.7% for dead leaves (Fig. 3). No significant difference in the estimated sampling rate $\hat{\tau}_c$ between the morning and the afternoon was detected whatever the type of vegetation or the number of successive samplings considered. Moreover, there was no significant difference between dead leaves and shrubs sampled in the morning (see Table 2 and Fig. 4). However, in the afternoon, the sampling rate of shrub sites was significantly lower than that of dead leaf sites sampled in the morning or in the afternoon (for $k=4$, the median of $\hat{\tau}$ posterior distribution, equal to 34.1% for shrubs in the afternoon versus 63.1% and 58.3% for dead leaves in the morning and in the afternoon respectively). The variability of the sampling rate $\hat{\sigma}_c^2$ due to the site observed decreased when the number of successive samplings considered in the model increased (Table 2). However, the variability was lower in dead leaves than in shrubs (Table 2).

4. Discussion

The hierarchical Bayesian approach proposed to estimate population size N_0 and sampling rate τ for a removal sampling design produces performant estimators i.e. estimation with small quadratic error. This method of estimation must be combined with an appropriate sampling design, which guarantees the main model assumption of a closed population.

The Bayesian framework presents major advantages. The first concerns the ability to take into account prior knowledge through a prior distribution on parameters (Gelman et al., 2004). The second major advantage lies in the construction of a complex model for a better understanding of the biological system observed. By constructing hierarchical models, one structure of dependence among observations can be specified and observable and/or unobservable

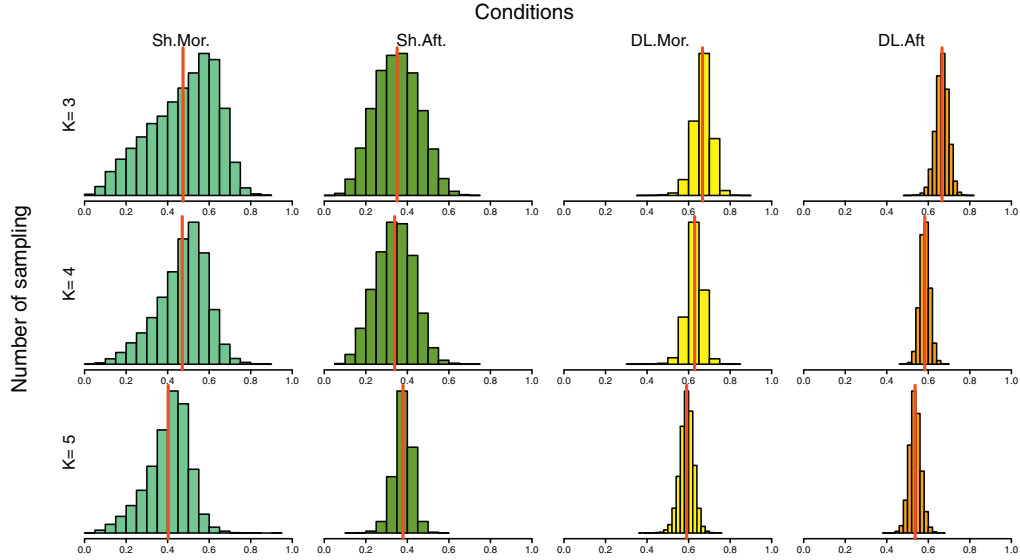


Fig. 3. Posterior distribution of sampling rate $\hat{\tau}$ according to conditions for HBM2 considering the 3, 4 and 5 first samplings for the 4 conditions: shrub sampled in the morning (Sh. Mor.), shrub sampled in the afternoon (Sh. Aft.), dead leaves sampled in the morning (DL Mor.), dead leaves sampled in the afternoon (DL Aft.). The red line represent the mean of posterior distribution of sampling rate $\hat{\tau}$.

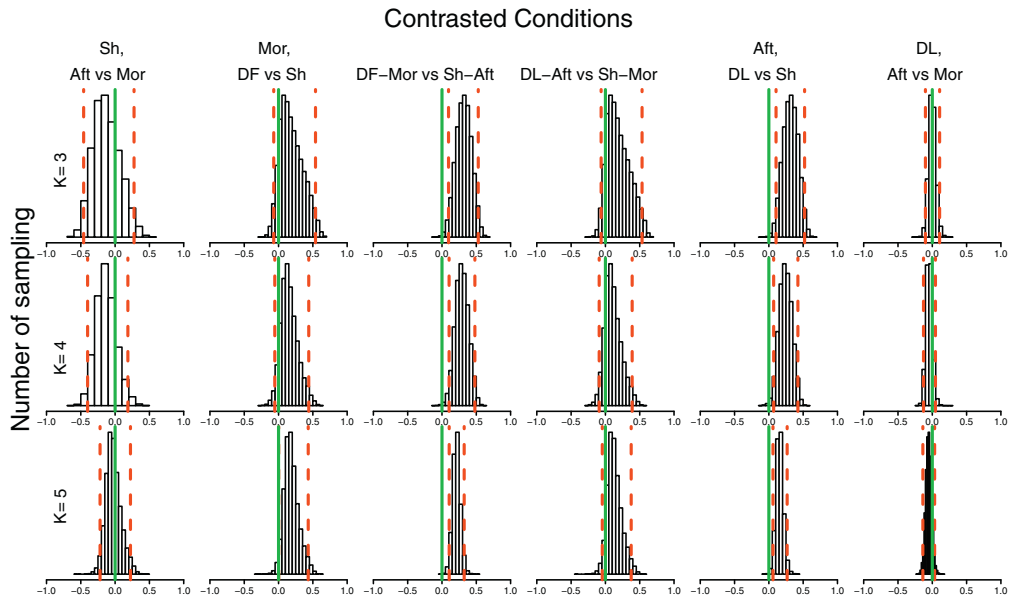


Fig. 4. Posterior distribution of contrasts of sampling rate estimation between conditions for HBM2 considering the 3, 4 and 5 first samplings. Comparison of sampling rate posterior distribution between shrub sampled in the morning versus shrub sampled in the afternoon (Sh., Aft. vs Mor.), dead leaves sampled in the morning and shrubs sampled in the morning (Mor., DL vs Sh.), dead leaves sampled in the morning and shrubs sampled in the afternoon (DF Mor. vs Sh. Aft.), dead leaves sampled in the afternoon and shrubs sampled in the morning (DL Aft. vs Sh. Mor.), dead leaves sampled in the afternoon and shrubs sampled in the afternoon (Aft., DL vs Sh.), dead leaves sampled in the afternoon and dead leaves sampled in the morning (DL, Aft. vs Mor.). Red dotted lines represent the credible interval at 95%. The green line marks the 0 value. If the green line was outside the interval, then we concluded that there was a significant difference of sampling rate of contrasted conditions.

Table 1

Ratio of the Root of Mean Square Error (RRMSE) values of the population size N_0 and the sampling rate τ estimates obtained by the Hayne method \hat{N}_{0Hayne} and HBM1 estimators for the simulated scenario $N_0 = 500$, $\tau = 0.1, 0.3, 0.5, 0.7$. \hat{N}_{0moy} and \hat{N}_{0med} correspond to the RRMSE of estimator based on the mean of posterior distribution and the median of posterior distribution respectively.

N_0	τ	k	\hat{N}_{0Hayne}	\hat{N}_{0moy}	\hat{N}_{0med}	$\hat{\tau}_{Hayne}$	$\hat{\tau}_{moy}$	$\hat{\tau}_{med}$
500	0.1	2	–	0.214	0.296	–	0.843	0.766
500	0.1	3	2.37×10^{13}	0.214	0.250	0.956	0.467	0.453
500	0.1	4	5.980	0.235	0.247	0.607	0.338	0.350
500	0.1	5	2.844	0.255	0.256	0.457	0.292	0.311
500	0.1	6	0.584	0.246	0.237	0.344	0.250	0.269
500	0.1	7	4.773	0.228	0.214	0.282	0.232	0.246
500	0.1	8	0.279	0.206	0.189	0.237	0.212	0.221
500	0.1	9	0.160	0.177	0.159	0.204	0.192	0.197
500	0.1	10	0.127	0.150	0.134	0.181	0.174	0.176
500	0.3	2	–	0.251	0.239	–	0.216	0.237
500	0.3	3	0.136	0.160	0.143	0.164	0.162	0.165
500	0.3	4	0.073	0.084	0.076	0.116	0.115	0.114
500	0.3	5	0.048	0.049	0.046	0.092	0.087	0.087
500	0.3	6	0.034	0.032	0.031	0.078	0.070	0.070
500	0.3	7	0.026	0.023	0.022	0.069	0.062	0.062
500	0.3	8	0.021	0.017	0.017	0.064	0.056	0.056
500	0.3	9	0.017	0.013	0.013	0.059	0.051	0.051
500	0.3	10	0.014	0.010	0.010	0.056	0.048	0.048
500	0.5	2	–	0.111	0.098	–	0.121	0.120
500	0.5	3	0.038	0.038	0.036	0.072	0.068	0.067
500	0.5	4	0.022	0.019	0.018	0.057	0.051	0.051
500	0.5	5	0.015	0.011	0.011	0.050	0.044	0.044
500	0.5	6	0.011	0.007	0.007	0.046	0.039	0.039
500	0.5	7	0.008	0.004	0.004	0.044	0.036	0.036
500	0.5	8	0.007	0.003	0.003	0.043	0.035	0.035
500	0.5	9	0.006	0.002	0.002	0.042	0.034	0.034
500	0.5	10	0.005	0.001	0.002	0.041	0.033	0.033
500	0.7	2	–	0.030	0.028	–	0.049	0.049
500	0.7	3	0.012	0.010	0.010	0.036	0.034	0.034
500	0.7	4	0.007	0.005	0.005	0.032	0.028	0.028
500	0.7	5	0.005	0.002	0.002	0.031	0.026	0.026
500	0.7	6	0.004	0.001	0.001	0.030	0.025	0.025
500	0.7	7	0.003	0.001	0.001	0.029	0.024	0.024
500	0.7	8	0.003	0.000	0.000	0.029	0.024	0.024
500	0.7	9	0.002	0.000	0.000	0.028	0.024	0.024
500	0.7	10	0.002	0.000	0.000	0.028	0.024	0.024

“–” for not estimated scenario by Hayne method.

parameters can be introduced (Royle and Dorazio, 2006; Gelman and Hill, 2006; Cressie et al., 2009).

The simulation study demonstrates that the performance of the Bayesian estimator is positively linked to the level of the sampling rate τ . When the sampling rate is equal to 10%, neither method performs well. This result is consistent with those of White et al. (1982) and Otis et al. (1978). When τ is higher than 10%, the model can give a relatively good quality estimate with only a few successive

samplings. This result allows the model's hypothesis of a closed population to be maintained. Indeed, during the few successive samplings needed to obtain a good estimation, the probability of a modification in the population size is low. This result also allows the cost associated with carrying out additional samplings to be reduced.

The hypothesis of a closed population assumes that there is no immigration or emigration and, therefore, that the number of

Table 2

Mean, median and quantiles at 2.5% and 97.5% of posterior distributions of the sampling rate τ_c and the variance parameter σ_c^2 of the random fluctuation of τ estimated by the hierarchical Bayesian model (HBM2) considering the 3, 4 and 5 first successive samplings for the 4 conditions: shrub sampled in the morning (Sh. Mor.), shrub sampled in the afternoon (Sh. Aft.), dead leaves sampled in the morning (DL Mor.), dead leaves sampled in the afternoon (DL Aft.).

k first sampling	Condition	$\hat{\tau}_c$					$\hat{\sigma}_c^2$				
		Mean	50%	SD	2.5%	97.5%	Mean	50%	SD	2.5%	97.5%
$k = 3$	Sh. Mor.	0.474	0.501	0.159	0.137	0.716	1.491	1.509	0.964	0.040	2.948
	Sh. Aft.	0.351	0.350	0.105	0.154	0.552	2.030	2.182	0.710	0.499	2.964
	DL Mor.	0.667	0.670	0.042	0.576	0.744	0.370	0.286	0.308	0.057	1.201
	DL Aft.	0.666	0.667	0.031	0.603	0.724	0.179	0.147	0.136	0.022	0.533
$k = 4$	Sh. Mor.	0.470	0.488	0.119	0.199	0.667	1.195	0.987	0.887	0.053	2.896
	Sh. Aft.	0.339	0.341	0.090	0.165	0.508	1.679	1.676	0.756	0.392	2.934
	DL Mor.	0.629	0.631	0.038	0.548	0.697	0.273	0.212	0.234	0.035	0.873
	DL Aft.	0.583	0.583	0.024	0.537	0.630	0.058	0.042	0.057	0.002	0.209
$k = 5$	Sh. Mor.	0.403	0.418	0.102	0.165	0.573	0.936	0.616	0.858	0.024	2.849
	Sh. Aft.	0.379	0.382	0.044	0.283	0.458	0.333	0.239	0.321	0.046	1.232
	DL Mor.	0.591	0.592	0.034	0.521	0.654	0.208	0.163	0.177	0.022	0.657
	DL Aft.	0.536	0.536	0.028	0.484	0.593	0.137	0.113	0.102	0.017	0.398

individuals in the population is constant over time (Otis et al., 1978). For a non-closed population, if there is immigration, the initial population size is overestimated and the sampling rate is underestimated. In contrast, if there is emigration, the initial population size is underestimated and the sampling rate is overestimated.

For ticks, immigration and emigration are the result of horizontal and/or vertical movements. To control horizontal movements, the observed site is usually closed. However, vertical movements are harder to control. If the interval between successive samplings is too long, active ticks, present on the ground, could “immigrate” (Sonenshine, 1994) into the host-seeking population, i.e. those waiting for a host on the top of the vegetation. Therefore, a short interval between successive samplings (2 min 30 s) was chosen to ensure that the population remained closed. Our estimation of sampling rate $\hat{\tau}$ consequently represents the sampling rate of host-seeking ticks only, which is equivalent to the estimation of the efficiency of the drag sampling method.

The sampling rate $\hat{\tau}$ was estimated to be between 33.9% and 47.4% for shrubs and 53.6% and 66.7% for dead leaves. The significant difference in the estimated sampling rates between shrubs and dead leaves could be explained by the level of contact between the drag and the layer of vegetation, which is higher in dead leaves than in shrubs (Milne, 1943; Ginsberg and Ewing, 1989; Petry et al., 2010).

Our estimations of sampling rates (around 50%) are higher than previous studies (lower than 10%) (Sonenshine, 1994; Daniels et al., 2000; Talleklint-Eisen and Lane, 2000). In the only study that used a RS approach (Talleklint-Eisen and Lane, 2000), the host-seeking tick population was not closed because the interval between successive samplings was so long (one day). Indeed, non host-seeking ticks had time to become host-seeking. Their sampling rate thus was in fact the joint probability of capturing the host-seeking ticks τ and the probability for one tick to be host-seeking when it was active.

Moreover, when the interval between successive samplings is too long, sampling conditions (such as temperature, humidity) may change from one sampling to the next. Consequently, the size of the host-seeking population could be modified and this new population could respond differently to the stimuli created by the drag. If this is so, the sampling rate would no longer be constant across successive samplings. An increase of the interval between samplings could hence lead to confusion between the variation of detection probability and the variation of population size. To estimate the abundance of the active tick population on a given site, RS with short intervals should be repeated every day during the entire active period to estimate both the sampling rate and the size of the host-seeking population. The size of the active tick population during the active period and the proportion of questing ticks each day could then be deduced.

The aim of our study was to analyse the performance of estimators and to study the spatial variations of detection probabilities. Spatial variations of the detection probability were taken into account in the model through a random effect. However, ticks are notoriously spatially aggregated at different scales (Boulinier et al., 1996; Elston et al., 2001). To analyse the abundance of ticks and to extrapolate it to a larger area, both, the variations of detection probabilities and the spatial aggregation of ticks should be taken into account. An optimal sampling design must balance the need to perform many successive samplings on each of the sites to estimate the detection probability with the need to explore many sites to analyse the spatial distribution of the population. The temporal variation of the detection probability due to the rank of sampling could also be considered. This requires a large increase in the number of observed sites. However, special attention should be paid to possible confusion between spatial and temporal variation when both variations are studied at the same time.

In conclusion, to determine the minimal number of successive samplings to be conducted, one needs to balance the cost of sampling (labour intensity, time, workforce) against the number needed to adequately model hypotheses and to achieve a good precision of estimation: too many successive samplings could lead to a non-closed population while an insufficient number of successive samplings could lead to poor estimates.

Acknowledgements

This study was partially funded by EU grant FP7-261504 EDENext and is catalogued by the EDENext Steering Committee as EDENext144 (<http://www.edenext.eu>). The contents of this publication are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission. We thank the «Tiques et Maladies» Tiques group of the «Réseau Ecologie des Interactions Durables» for discussion and support. We are thankful to the Office National des Forêts for allowing us to work in Sénart forest. Very sincere thanks to Elise Vaumourin for her help in data acquisition. Thanks to Grace Delobel for proofreading the text in English.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.ecolmodel.2013.12.004>. These data include Google maps of the most important areas described in this article.

References

- Allredge, M.W., Pollock, K.H., Simons, T.R., Collazo, J.A., Shriner, S.A., 2007. Time-of-detection method for estimating abundance from point-count surveys. *Auk* 124, 653–664.
- Anderson, D.R., 2001. The need to get the basics right in wildlife field studies. *Wildlife Society Bulletin* 29, 1294–1297.
- Anderson, R.M., 1991. Populations and infectious diseases: ecology or epidemiology? *Journal of Animal Ecology* 60, 1–50.
- Boulinier, T., Ives, A.R., Danchin, E., 1996. Measuring aggregation of parasites at different host population levels. *Parasitology* 112, 581–587.
- Clark, J.S., Gelfand, A.E., 2006. A future for models and data in environmental science. *Trends in Ecology & Evolution* 21, 375–380.
- Cressie, N., Calder, C.A., Clark, J.S., Hoef, J.M.V., Wikle, C.K., 2009. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *Ecological Applications* 19, 553–570.
- Daniels, T.J., Falco, R.C., Fish, D., 2000. Estimating population size and drag sampling efficiency for the blacklegged tick (Acari: Ixodidae). *Journal of Medical Entomology* 37, 357–363.
- Dodd, C.K., Dorazio, R.M., 2004. Using counts to simultaneously estimate abundance and detection probabilities in a salamander community. *Herpetologica* 60, 468–478.
- Elston, D.A., Moss, R., Boulinier, T., Arrowsmith, C., Lambin, X., 2001. Analysis of aggregation, a worked example: numbers of ticks on red grouse chicks. *Parasitology* 122, 563–569.
- Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. *Bayesian Data Analysis*. Chapman & Hall/CRC, London.
- Gelman, A., Hill, J., 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*, 1st ed. Cambridge University Press, Cambridge, England.
- Gelman, A., Shirley, K., 2011. Inference from simulations and monitoring convergence. In: *Handbook of Markov Chain Monte Carlo*. CRC Press, Boca Raton, FL, pp. 163–174.
- Ginsberg, H.S., Ewing, C.P., 1989. Comparison of flagging, walking, trapping, and collecting from hosts as sampling methods for northern deer ticks, *Ixodes dammini*, and lone-star ticks, *Amblyomma americanum* (Acari: Ixodidae). *Experimental and Applied Acarology* 7, 313–322.
- Hayne, D.W., 1949. Two methods for estimating population from trapping records. *Journal of Mammalogy* 30, 399–411.
- Kery, M., 2004. Extinction rate estimates for plant populations in revisitation studies: importance of detectability. *Conservation Biology* 18, 570–574.
- Kery, M., Royle, J.A., 2010. Hierarchical modelling and estimation of abundance and population trends in metapopulation designs. *Journal of Animal Ecology* 79, 453–461.
- Lehmann, E., Casella, G., 1998. *Theory of Point Estimation*, 2nd ed. Springer, New York.
- MacKenzie, D.I., Kendall, W.L., 2002. How should detection probability be incorporated into estimates of relative abundance? *Ecology* 83, 2387–2393.

- Milne, A., 1943. The comparison of sheep-tick populations (*Ixodes ricinus* L.). *Annals of Applied Biology* 30, 240–250.
- Moran, P.A.P., 1951. A mathematical theory of animal trapping. *Biometrika* 38, 307–311.
- Otis, D.L., Burnham, K.P., White, G.C., Anderson, D.R., 1978. Statistical inference from capture data on closed animal populations. *Wildlife Monographs* 62, 1–135.
- Parola, P., Raoult, D., 2001. Ticks and tickborne bacterial diseases in humans: an emerging infectious threat. *Clinical Infectious Diseases* 32, 897–928.
- Pellet, J., Schmidt, B.R., 2005. Monitoring distributions using call surveys: estimating site occupancy, detection probabilities and inferring absence. *Biological Conservation* 123, 27–35.
- Petry, W., Foré, S., Fielden, L., Kim, H.J., 2010. A quantitative comparison of two sample methods for collecting *Amblyomma americanum* and *Dermacentor variabilis* (Acari: Ixodidae) in Missouri. *Experimental and Applied Acarology* 52, 427–438.
- Pollock, K.H., Nichols, J.D., Brownie, C., Hines, J.E., 1990. Statistical inference for capture–recapture experiments. *Wildlife Monographs* 107, 1–97.
- Pollock, K.H., Nichols, J.D., Simons, T.R., Farnsworth, G.L., Bailey, L.L., Sauer, J.R., 2002. Large scale wildlife monitoring studies: statistical methods for design and analysis. *Environmetrics* 13, 105–119.
- Royle, J.A., Dorazio, R.M., 2006. Hierarchical models of animal abundance and occurrence. *Journal of Agricultural, Biological, and Environmental Statistics* 11, 249–263.
- Sonenshine, D.E., 1994. *Ecological Dynamics of Tick-Borne Zoonoses*. Oxford University Press.
- Talleklint-Eisen, L., Lane, R.S., 2000. Spatial and temporal variation in the density of *Ixodes pacificus* (Acari: Ixodidae) nymphs. *Environmental Entomology* 29, 272–280.
- Thompson, W.L., White, G.C., Gowan, C., 1998. *Monitoring Vertebrate Populations*. San Diego Academic Press.
- Thulasiraman, K., Swamy, M., 1992. *Graphs: Theory and Algorithms*. Wiley & Sons, Incorporated, New York, pp. 118 (Chapter 5.7 Acycli).
- Vassallo, M., Pichon, B., Cabaret, J., Figureau, C.U., Perez-Eid, C., 2000. Methodology for sampling questing nymphs of *Ixodes ricinus* (Acari: Ixodidae), the principal vector of lyme disease in Europe. *Journal of Medical Entomology* 37, 335–339.
- White, G.C., Anderson, D.R., Burnham, K.P., Otis, D.L., 1982. *Capture–Recapture and Removal Methods for Sampling Closed Populations*. Los Alamos National Laboratory, LA-8787-NERP.
- Williams, B.K., Nichols, J.D., Conroy, M.J., 2002. *Analysis and Management of Animal Populations*, 1st ed. Academic Press, San Diego, USA & London, UK.
- Zippin, C., 1956. An evaluation of the removal method of estimating animal populations. *Biometrics* 12, 163–189.

```
model
{
  # Specify observation model and prior for local parameters
  for (s in 1:n){
    NO[s] ~ dcat(pn[])
    tau[s] ~ dbeta(1,1)
    x[1+(k*(s-1))] ~ dbin(tau[s], NO[s])
    z[1+(k*(s-1))] <- NO[s]-x[(1+k*(s-1))]
    for (l in 2:k){
      x[l+(k*(s-1))] ~ dbin(tau[s], z[(l-1)+(k*(s-1))])
      z[l+(k*(s-1))] <- NO[s]-sum(x[(1+k*(s-1)):(l+k*(s-1))])
    }
  }
}
```

Supp MAT: BUGS code for fitting the Hierarchical Bayesian Model 1 (HBM1) of removal sampling data for *I. ricinus* counts.

```
model{
  # Specify priors for metapopulation-level parameters
  for (c in 1:nb_c){
    tau_c[c] ~ dbeta(1,1)
    sigma2_c[c] ~ dunif(0,3)
  }
  #
  # Specify observation model and prior for local parameters
  for (s in 1:n){
    NO[s] ~ dcat(pn[ ])
    epsilon[s] ~ dnorm(0, sigma2_epsi[s])
    #
    sigma2_epsi[s] <- 1/(sigma2_c[f_c[s]])
    logit_tau_c[s] <- log((tau_c[f_c[s]])/(1-(tau_c[f_c[s]])))
    logit_tau_s[s] <- logit_tau_c[s] + epsilon[s]
    tau[s] <- exp(logit_tau_s[s])/(1+(exp(logit_tau_s[s])))
    #
    x[1+(k*(s-1))] ~ dbin(tau[s], NO[s])
    z[1+(k*(s-1))] <- NO[s]-x[(1+k*(s-1))]
    for(l in 2:k){
      x[l+(k*(s-1))] ~ dbin(tau[s], z[(l-1)+(k*(s-1))])
      z[l+(k*(s-1))] <- NO[s]-sum(x[(1+k*(s-1)):(l+k*(s-1))])
    }
  }
}
```

Supp MAT: BUGS code for fitting the Hierarchical Bayesian Model 2 (HBM2) of removal sampling data for *I. ricinus* counts.

N_0	τ	k	\hat{N}_{0Hayne}	\hat{N}_{0moy}	\hat{N}_{0med}	$\hat{\tau}_{Hayne}$	$\hat{\tau}_{moy}$	$\hat{\tau}_{med}$
50	0.1	2	-	2.979	1.093	-	1.269	1.224
50	0.1	3	Inf	3.227	1.432	3.646	0.977	1.038
50	0.1	4	Inf	3.318	1.596	2.051	0.826	0.924
50	0.1	5	Inf	3.355	1.688	1.523	0.748	0.858
50	0.1	6	9.74×10^{13}	3.324	1.705	1.158	0.674	0.784
50	0.1	7	8.12×10^{13}	3.197	1.654	0.951	0.611	0.712
50	0.1	8	12.617	3.041	1.543	0.798	0.574	0.667
50	0.1	9	6.02×10^{14}	2.829	1.423	0.697	0.544	0.623
50	0.1	10	1.42×10^{14}	2.617	1.303	0.618	0.525	0.589
50	0.3	2	-	3.510	1.861	-	0.483	0.623
50	0.3	3	8.58×10^{13}	2.896	1.495	0.537	0.472	0.548
50	0.3	4	8.682	2.036	0.977	0.382	0.408	0.442
50	0.3	5	0.359	1.215	0.519	0.307	0.331	0.341
50	0.3	6	0.149	0.659	0.260	0.265	0.273	0.273
50	0.3	7	0.095	0.310	0.127	0.236	0.222	0.221
50	0.3	8	0.071	0.142	0.073	0.216	0.193	0.192
50	0.3	9	0.056	0.070	0.049	0.203	0.175	0.174
50	0.3	10	0.046	0.045	0.036	0.193	0.162	0.161
50	0.5	2	-	2.466	1.194	-	0.411	0.453
50	0.5	3	0.299	0.923	0.397	0.221	0.267	0.266
50	0.5	4	0.086	0.309	0.133	0.178	0.183	0.181
50	0.5	5	0.049	0.070	0.045	0.155	0.141	0.140
50	0.5	6	0.035	0.032	0.026	0.144	0.124	0.124
50	0.5	7	0.027	0.018	0.017	0.137	0.115	0.115
50	0.5	8	0.022	0.011	0.011	0.132	0.108	0.108
50	0.5	9	0.018	0.007	0.006	0.129	0.104	0.105
50	0.5	10	0.015	0.005	0.004	0.127	0.101	0.101
50	0.7	2	-	0.783	0.316	-	0.232	0.220
50	0.7	3	0.042	0.064	0.042	0.113	0.115	0.111
50	0.7	4	0.024	0.020	0.017	0.100	0.090	0.090
50	0.7	5	0.016	0.008	0.007	0.095	0.081	0.081
50	0.7	6	0.012	0.004	0.004	0.092	0.077	0.078
50	0.7	7	0.010	0.002	0.002	0.091	0.075	0.076
50	0.7	8	0.008	0.001	0.001	0.089	0.075	0.076
50	0.7	9	0.007	0.000	0.000	0.089	0.075	0.076
50	0.7	10	0.006	0.000	0.000	0.088	0.074	0.075

Table A. 1: Ratio of the Root of Mean Square Error (RRMSE) values of the population size N_0 and the sampling rate τ estimates obtained by the Hayne method \hat{N}_{0Hayne} and HBM1 estimators for the simulated scenario $N_0 = 50$, $\tau = 0.1, 0.3, 0.5, 0.7$. \hat{N}_{0moy} and \hat{N}_{0med} correspond to the RRMSE of estimator based on the mean of posterior distribution and the median of posterior distribution respectively. "-" for not estimated scenario by Hayne method.

N_0	τ	k	\hat{N}_{0Hayne}	\hat{N}_{0moy}	\hat{N}_{0med}	$\hat{\tau}_{Hayne}$	$\hat{\tau}_{moy}$	$\hat{\tau}_{med}$
100	0.1	2	-	1.510	0.718	-	1.039	1.008
100	0.1	3	Inf	1.685	0.945	2.269	0.843	0.906
100	0.1	4	8.69×10^{13}	1.787	1.080	1.420	0.603	0.691
100	0.1	5	1.54×10^{14}	1.805	1.136	1.026	0.535	0.628
100	0.1	6	4.46×10^{13}	1.748	1.113	0.809	0.535	0.615
100	0.1	7	4.23×10^{13}	1.625	1.023	0.646	0.485	0.554
100	0.1	8	6.64×10^{13}	1.509	0.953	0.557	0.465	0.520
100	0.1	9	3.928	1.361	0.845	0.488	0.438	0.481
100	0.1	10	1.198	1.186	0.722	0.426	0.404	0.435
100	0.3	2	-	1.815	1.169	-	0.445	0.534
100	0.3	3	21.849	1.278	0.781	0.373	0.374	0.407
100	0.3	4	2.250	0.673	0.379	0.260	0.283	0.287
100	0.3	5	0.134	0.291	0.172	0.206	0.210	0.208
100	0.3	6	0.083	0.134	0.090	0.177	0.171	0.170
100	0.3	7	0.060	0.067	0.056	0.157	0.143	0.142
100	0.3	8	0.047	0.046	0.041	0.145	0.128	0.128
100	0.3	9	0.038	0.034	0.031	0.135	0.116	0.116
100	0.3	10	0.032	0.025	0.024	0.128	0.108	0.108
100	0.5	2	-	0.985	0.554	-	0.315	0.324
100	0.5	3	0.104	0.225	0.138	0.158	0.172	0.169
100	0.5	4	0.053	0.061	0.051	0.124	0.117	0.116
100	0.5	5	0.034	0.028	0.026	0.107	0.093	0.092
100	0.5	6	0.024	0.017	0.016	0.099	0.084	0.084
100	0.5	7	0.019	0.011	0.011	0.094	0.078	0.078
100	0.5	8	0.015	0.007	0.008	0.091	0.075	0.075
100	0.5	9	0.013	0.005	0.005	0.089	0.072	0.072
100	0.5	10	0.011	0.003	0.003	0.087	0.071	0.071
100	0.7	2	-	0.143	0.091	-	0.129	0.123
100	0.7	3	0.027	0.027	0.024	0.078	0.075	0.074
100	0.7	4	0.016	0.011	0.011	0.070	0.061	0.061
100	0.7	5	0.011	0.005	0.006	0.066	0.057	0.057
100	0.7	6	0.009	0.003	0.002	0.064	0.055	0.055
100	0.7	7	0.007	0.001	0.001	0.063	0.054	0.054
100	0.7	8	0.006	0.000	0.000	0.063	0.053	0.054
100	0.7	9	0.005	0.000	0.000	0.062	0.053	0.053
100	0.7	10	0.004	0.000	0.000	0.062	0.053	0.053

Table A. 2: Ratio of the Root of Mean Square Error (RRMSE) values of the population size N_0 and the sampling rate τ estimates obtained by the Hayne method \hat{N}_{0Hayne} and HBM1 estimators for the simulated³ scenario $N_0 = 100$, $\tau = 0.1, 0.3, 0.5, 0.7$. \hat{N}_{0moy} and \hat{N}_{0med} correspond to the RRMSE of estimator based on the mean of posterior distribution and the median of posterior distribution respectively. "-" for not estimated scenario by Hayne method.

Influence de l'*a priori* sur les estimateurs Bayésiens

Pour estimer l'abondance ou le taux d'échantillonnage en l'absence d'information *a priori*, nous pouvons soit utiliser des méthodes fréquentistes, soit des méthodes bayésiennes avec des *a priori* non informatifs. Nous avons vu dans les chapitres précédents que le maximum de vraisemblance était parfois instable et donnait des estimateurs divergents de l'abondance. Nous avons essayé d'utiliser des *a priori* non informatifs, par exemple des lois plates sur N_0 et τ , ou des *a priori* vagues sur N_0 comme des lois uniformes sur un intervalle très grand. Nous avons constaté des résultats aberrants sur les estimations ainsi que des comportements particuliers sur les chaînes de Markov. En effet, les chaînes de Markov utilisées dans les méthodes MCMC, pouvaient diverger ou avoir des problèmes de mélangeance¹ faisant apparaître plusieurs modes. Nous avons cherché à comprendre ces phénomènes.

Dans ce chapitre, nous étudions de manière théorique les propriétés du modèle associé au « removal sampling » (RS). Nous exposons ici les principaux résultats, leurs démonstrations se trouvant dans la proposition d'article en fin de chapitre. Nous avons également mené des simulations afin, d'une part, d'illustrer les résultats théoriques et, d'autre part, de fournir des recommandations aux utilisateurs.

Dans la plupart des modèles statistiques, la vraisemblance tend vers 0 sur les bords du domaine étudié. Ceci entraîne la convergence de l'estimateur du maximum de vraisemblance ainsi qu'une certaine stabilité des estimateurs bayésiens avec des *a priori* vagues. Ce n'est pas le cas pour le RS. En fait, le modèle RS a pour modèle limite un modèle de Poisson iid² quand N_0 tend vers l'infini et donc τ tend vers 0 (cf. Fig. 5.1 et Fig. 5.2). Cela signifie que, lorsque N_0 est grand, la vraisemblance du modèle associé au RS peut être approchée par la vraisemblance d'un modèle de Poisson iid (cf. Proposition 2.1. de l'article).

Le problème est que ce modèle limite n'est pas identifiable en N_0 et τ , puisqu'il

1. capacité de la chaîne de Markov à visiter l'ensemble du domaine de définition du paramètre et à sortir des minimums locaux

2. indépendant et identiquement distribué

dépend uniquement du paramètre $\lambda = N_0 \times \tau$. En effet, pour une même valeur de λ , il existe une multitude de combinaisons de valeurs possibles pour N_0 et τ dont nous ne pouvons extraire celle qui a généré les données observées. La conséquence est que la vraisemblance ne tend pas vers 0 quand N_0 tend vers $+\infty$ et $N_0 \times \tau$ tend vers une valeur λ . La figure 5.1 montre un exemple de vraisemblance $L(N_0, \tau)$ associée au modèle de « removal sampling » avec les paramètres (N_0, τ) , alors que la figure 5.3 montre la même vraisemblance $L(N_0, \lambda)$ en reparamétrant le modèle en (N_0, λ) . A N_0 fixé grand, nous avons $L(N_0, \lambda) \simeq L_P(\lambda)$ où $L_P(\lambda)$ est la vraisemblance d'un modèle de Poisson de paramètre λ . Ce phénomène limite s'observe également sur la vraisemblance profilée $L(N_0, \tau)$ (cf. Fig. 5.4 et Proposition 3.2. de l'article). Nous pouvons noter qu'elle ne tend pas vers 0 quand N_0 tend vers l'infini, mais vers le maximum de vraisemblance dans un modèle de Poisson iid comme illustré en rouge sur la figure 5.5.

En ce qui concerne l'analyse bayésienne, l'impact de ce phénomène est encore plus important. En effet, si le poids de l'*a priori* est trop fort sur des valeurs de τ faibles et/ou des valeurs de N_0 élevées, alors le fait que la vraisemblance ne tende pas vers 0 peut amener à des *a posteriori* impropres³ ou des estimateurs divergents ou instables. Nous reprenons ici quelques résultats obtenus dans l'article :

Cas 1 : prenons pour *a priori* non informatif $\pi(N_0, \tau) \propto 1$ (*a priori* de Laplace ou flat prior), alors l'*a posteriori* est impropre et l'estimateur bayésien n'est pas défini.

Cas 2 : pour *a priori* $\pi(N_0, \tau) \propto \frac{1}{\tau(1-\tau)}$ (*a priori* de Haldane sur τ , cf. Fig. 5.6a) ou $\pi(N_0, \tau) \propto \frac{1}{\sqrt{\tau(1-\tau)}}$ (*a priori* de Jeffreys sur τ , cf. Fig. 5.6b), alors ici également l'*a posteriori* est impropre. Cela s'explique par le fait que par rapport à l'*a priori* de Laplace, on met davantage de poids autour de 0 et de 1 pour τ .

Cas 3 : on diminue le poids de l'*a priori* pour les valeurs élevées de N_0 en prenant l'*a priori* impropre $\pi(N_0, \tau) \propto \frac{1}{N_0}$ (cf. Fig. 5.6c), on obtient bien un *a posteriori* propre, mais l'estimateur bayésien de N_0 diverge.

Cas 4 : on diminue davantage le poids de l'*a priori* pour les valeurs élevées de N_0 en prenant l'*a priori* propre $\pi(N_0, \tau) \propto \frac{1}{N_0^2}$ (cf. Fig. 5.6d). L'*a priori* est alors propre et l'estimateur converge.

Cas 5 : plutôt que de diminuer le poids sur les N_0 grands, on peut garder un *a priori* de Laplace sur N_0 mais diminuer le poids sur les valeurs petites de τ en posant pour *a priori* une loi $\beta(a, b)$ avec $a > 1$. Dans ce cas, l'*a posteriori* est propre, et l'estimateur de N_0 diverge si et seulement si $b > 2$.

Cas 6 : une autre possibilité est d'approcher une loi pour lequel l'estimateur diverge par une loi pour lequel l'estimateur converge. Par exemple, on peut utiliser sur N_0 une loi uniforme sur $\{0, 1, \dots, A\}$ où A est très grand. On peut également

3. de masse infinie *i.e.* une loi *a priori* pour $\theta \in \Theta$ est impropre si $\int_{\Theta} \pi(\theta) d\theta = \infty$, et est propre si $\int_{\Theta} \pi(\theta) d\theta < \infty$. Par exemple, une loi uniforme sur \mathcal{R} est impropre.

approcher une loi $\beta(a, b)$ sur τ par une loi $\beta(a+\varepsilon, b+\varepsilon)$. Cependant, on montre que les estimateurs sont instables car ils sont très dépendants du choix des hyperparamètres A ou ε *i.e.* tendent vers $+\infty$ quand $A \rightarrow +\infty$ ou $\varepsilon \rightarrow 0$.

Ces résultats théoriques donnent des indications sur la façon de choisir la loi *a priori* en absence d'information *a priori*. En particulier, il permet d'éliminer certains types de lois. Afin d'affiner ces résultats, nous avons mené des simulations selon plusieurs scénarii.

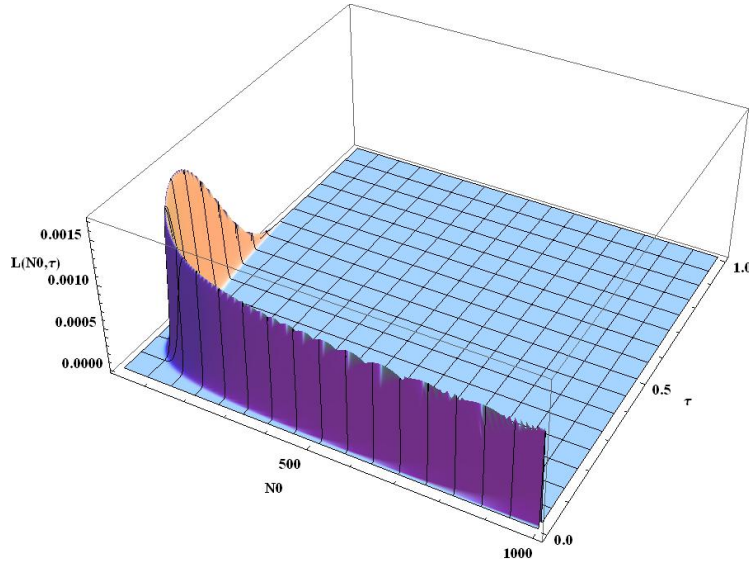


Fig. 5.1 : Vraisemblance $L(N_0, \tau)$ du modèle de « removal sampling » pour la séquence $X = \{13, 8, 10\}$.

5.1 Schéma de simulations et critère de comparaison des estimateurs

Pour étudier l'effet des distributions *a priori* sur le comportement des estimateurs, nous avons réalisé des simulations en considérant $N_0 = [50; 500]$ et $\tau = [0,1; 0,3; 0,5]$ pour 3 échantillonnages successifs. 50 jeux de données ont été simulés pour chaque scénario. Nous avons considéré plusieurs couples de loi *a priori* pour N_0 et τ (cf. Table 5.1).

La qualité d'un estimateur $\hat{\theta}$ du paramètre θ a été évaluée par le rapport entre la racine carré de l'Erreur Quadratique Moyenne⁴ (MSE) (Lehmann and Casella,

4. Mean Square Error en Anglais

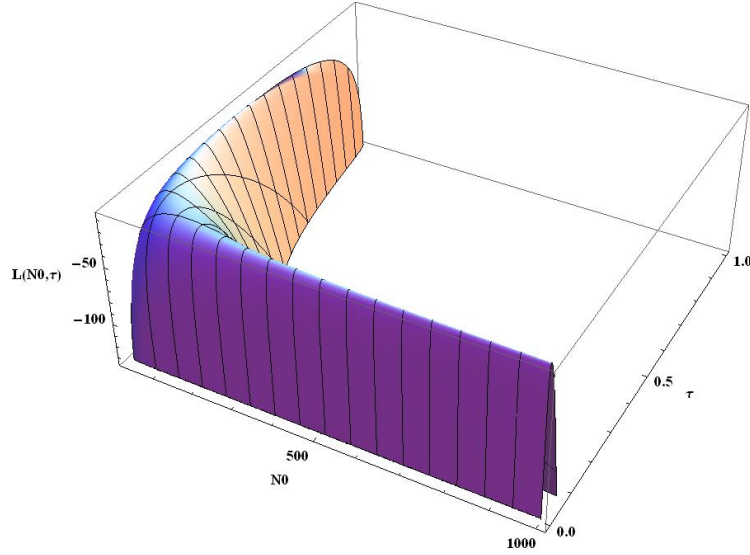


Fig. 5.2 : Log-Vraisemblance $L(N_0, \tau)$ du modèle de « removal sampling » pour la séquence $X = \{13, 8, 10\}$.

TABLE 5.1 : Couples de lois *a priori* pour N_0 et τ avec $A = [10^4; 10^5; 10^6; 10^7]$ et $B = [10^6; 10^7]$

Prior N_0	Prior τ	Cas
U sur $\{1, \dots, A\}$	$\beta(2, 2)$	5
U sur $\{1, \dots, B\}$	$\beta(2, 1, 10)$	6
U sur $\{1, \dots, B\}$	$\beta(3, 3)$	6
dflat	$\beta(2, 1, 10)$	6
dflat	$\beta(3, 3)$	6
$1/N_0^2$	$\beta(3, 3)$	4

1998) et la vraie valeur du paramètre θ notée $RRMSE$. Ce critère est basé sur le MSE dont les 2 composants représentent la variabilité de l'estimateur et le biais *i.e.* la distance entre la valeur estimée et la vraie valeur du paramètre :

$$MSE_{\hat{\theta}}(\theta) = \mathbb{E}((\hat{\theta} - \theta)^2) = (\text{biais}(\hat{\theta}))^2 + \text{var}(\hat{\theta})$$

avec,

$$\text{biais}(\hat{\theta}) = \mathbb{E}(\hat{\theta} - \theta).$$

La racine carré du MSE est souvent préférée car elle est du même ordre de grandeur que le paramètre. Pour avoir un critère comparable entre les valeurs des paramètres à estimer, nous avons considéré le RRMSE. L'estimateur ayant la plus petite valeur de RRMSE est considéré comme étant le meilleur estimateur du paramètre.

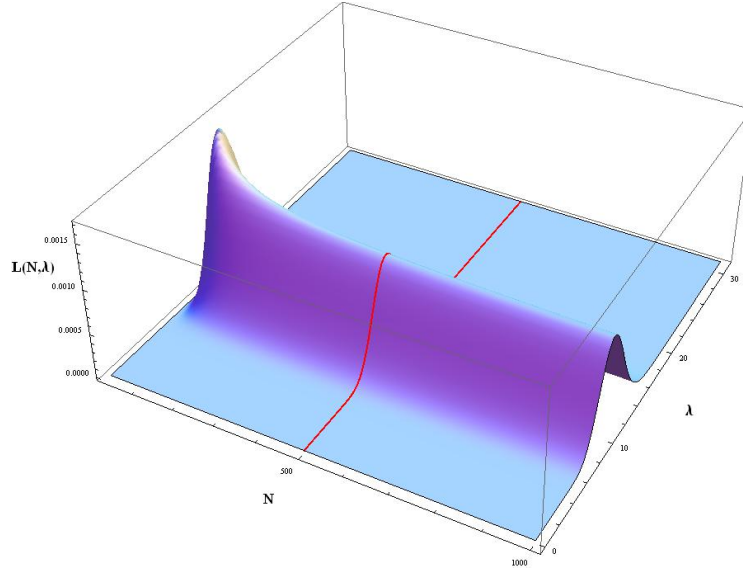


Fig. 5.3 : Vraisemblance $L(N_0, \lambda)$ du modèle de « removal sampling » pour la séquence $X = \{13, 8, 10\}$.

5.2 Résultats des simulations

5.2.1 Comportement des estimateurs dans le cas non convergent (cas 5)

Si on considère une loi *a priori* $\beta(2, 2)$ pour τ et une loi uniforme sur $[1, A]$ pour N_0 avec $A = \{10^4, 10^5, 10^6, 10^7\}$ alors les estimateurs de N_0 divergent (cf. Table 5.2).

Pour $N_0 = 50$ ou pour $N_0 = 500$ et $\tau = 0,1$, l'estimateur de N_0 basé sur la moyenne de la loi *a posteriori* noté $\hat{N}_{0_{moy}}$ tend vers l'infini lorsque la borne A de la loi uniforme *a priori* de N_0 tends vers l'infini (cf. Table 5.2). Par contre, $N_0 = 500$ et $\tau \geq 0,3$, l'estimateur $\hat{N}_{0_{moy}}$ ne semble plus être influencé par la borne A de la loi uniforme (cf. Table 5.2). Ce résultat paraît contradictoire avec les résultats théoriques présentés dans l'article suivant (cf. Proposition 4.7). Nous détaillons ici les raisons de ce résultat.

Lorsque l'on considère une loi *a priori* $\beta(1, 1)$ sur τ et une loi plate sur N_0 , alors pour N_0 est grand, la vraisemblance marginale $\pi(N_0|x)$ se comporte comme $\frac{K}{N_0^a}$ (cf. Lemme 4.2 de l'article suivant), ce qui donne $\pi(N_0|x) \approx K/N_0^a$ et l'espérance de la loi marginale vaut $\mathbb{E}(N_0|x) \approx \frac{K}{N_0^{a-1}}$.

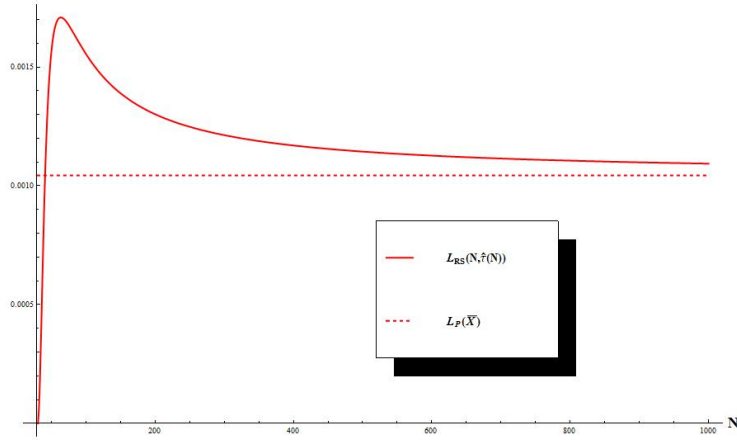


Fig. 5.4 : Vraisemblances Profilées $L(N_0, \lambda)$ et $L_P(\lambda)$ du modèle de « removal sampling » pour la séquence $X = \{13, 8, 10\}$.

On peut décomposer $\mathbb{E}(N_0|x)$ comme :

$$\mathbb{E}(N_0|x) \approx \sum_{i=\sum_{j=1}^k x_j}^A N_0 \pi(N_0|x) + \sum_{i=A+1}^{A+B} N_0 \pi(N_0|x),$$

où, A est choisi tel que la vraisemblance du « removal sampling » peut être approchée par la vraisemblance d'un modèle de Poisson et B représente la borne supérieure de la loi *a priori* uniforme de N_0 .

On pose, $D = \sum_{i=\sum_{j=1}^k x_j}^A N_0 \pi(N_0|x)$ et $F = \sum_{i=A+1}^{A+B} \frac{K}{N_0^{a-1}}$.

F dépend du paramètre a de la loi $\beta(a, b)$ de τ et de B , la borne supérieure de la loi uniforme de N_0 .

Si K est négligeable par rapport à D , alors le choix de B aura peu d'influence sur l'estimateur de N_0 basé sur l'espérance de la loi *a posteriori* sauf si B est extrêmement grand. C'est ce phénomène que l'on observe pour le scénario $N_0 = 500$ et $\tau \geq 0,3$ (cf. Table 5.2).

Prenons par exemple, la séquence $x = \{238, 135, 61\}$ simulée pour $N_0 = 500$ et $\tau = 0,5$. On considère une loi *a priori* $\beta(1, 1)$ pour τ et une loi uniforme sur $[1, B]$ pour N_0 , alors on a pour $A = 1000$:

- $K \approx 2,34.10^{-28}$
- $D \approx 0,07158$

Ici, K est négligeable par rapport à D ; par conséquent, pour $B < 10^{29}$, l'estimateur de N_0 basé sur l'espérance de la loi *a posteriori* ne sera pas influencé par le choix de B (c'est ce qui se passe pour les simulations du scénario $N_0 = 500$ et $\tau = 0,5$

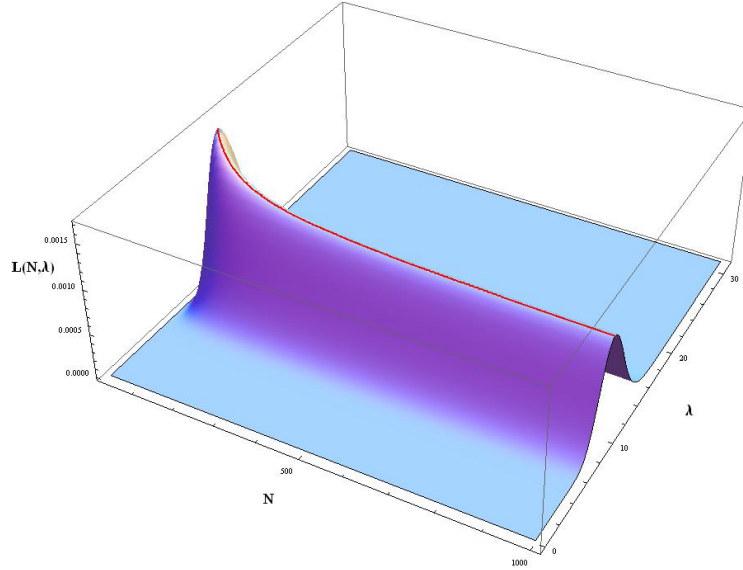


Fig. 5.5 : Vraisemblance $L(N_0, \lambda)$ et vraisemblance profilée (en rouge) du modèle de « removal sampling » pour la séquence $X = \{13, 8, 10\}$.

présentées dans la Table 5.2).

Si on considère la séquence $x = \{24, 17, 5\}$ simulée pour $N_0 = 50$ et $\tau = 0,5$ alors on a, pour $A = 1000$:

- $K \approx 7,81.10^{-6}$
- $D \approx 0,2131$

Ici, K n'est pas négligeable par rapport à D . Par conséquent l'estimateur de N_0 basé sur l'espérance de la loi marginale *a posteriori* sera sensible à la borne A (c'est ce qui se passe pour les simulations du scénario $N_0 = 50$ et $\tau = 0,5$ présentées dans la Table 5.2).

Concernant les estimateurs de N_0 basés sur la médiane noté $\hat{N}_{0_{\text{moy}}}$ et de τ ne semblent pas influencés par le choix de l'*a priori*. Globalement, les estimateurs basés sur la médiane semblent meilleurs que ceux basés sur la moyenne (valeurs de RRMSE inférieures).

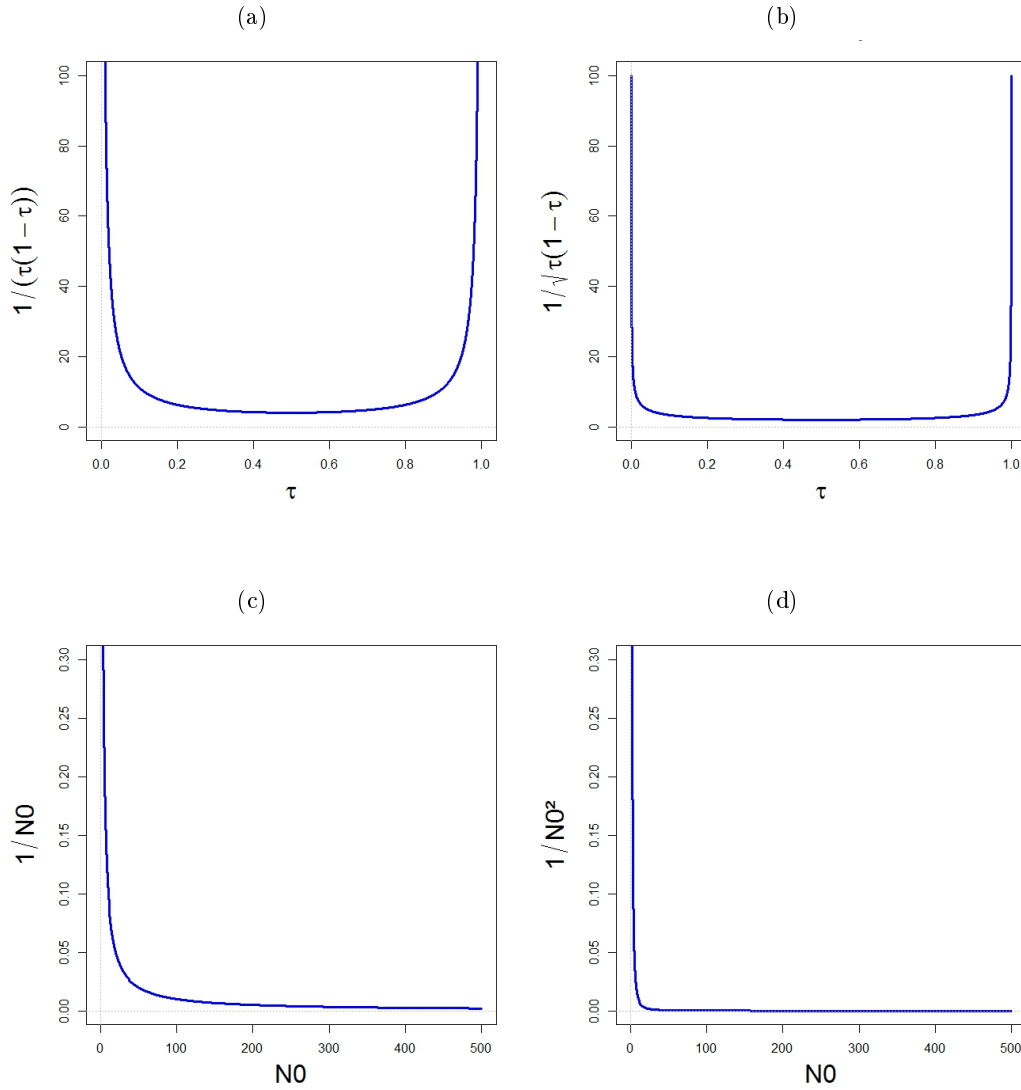


Fig. 5.6 : Distributions *a priori* pour τ : Haldane (a), Jeffreys (b) et distributions *a priori* pour N_0 : $1/N_0$ (c) et $1/N_0^2$ (d)

5.2.2 Comportement des estimateurs dans le cas convergent (cas 4 et 6)

- Pour $\tau = 0,1$, les estimateurs de
 - N_0 basés sur la moyenne de la loi *a posteriori* ($\hat{N}_{0_{moy}}$) sont meilleurs (RRMSE inférieurs) avec une loi *a priori* pour $\tau \sim \beta(3,3)$ par rapport à une loi $\beta(2,1,10)$ (cf. Table 5.3). Ils sont similaires entre une loi Uniforme et une

TABLE 5.2 : Comportement des estimateurs dans le cas non convergent avec une loi *a priori* $\beta(2, 2)$ pour τ et une loi uniforme sur $[1, A]$ pour N_0 avec $A = \{10^4, 10^5, 10^6, 10^7\}$. $\hat{N}_{0_{moy}}$ et $\hat{\tau}_{moy}$ représentent respectivement l'estimation de N_0 et de τ par la moyenne de la distribution *a posteriori* ; $\hat{N}_{0_{med}}$ et $\hat{\tau}_{med}$ représentent respectivement l'estimation de N_0 et de τ par la médiane de la distribution *a posteriori* ; RRMSE représente le ratio entre la racine carré de l'Erreur Quadratique Moyenne et la vraie valeur du paramètre.

N_0	τ	$\pi(N_0)$	$\hat{N}_0 = \hat{N}_{0_{moy}}$			$\hat{\tau} = \hat{\tau}_{moy}$			$\hat{N}_0 = \hat{N}_{med}$			$\hat{\tau} = \hat{\tau}_{med}$		
			$\mathbb{E}(\hat{N}_0)$	$V(\hat{N}_0)$	RRMSE	$\mathbb{E}(\hat{\tau})$	$V(\hat{\tau})$	RRMSE	$\mathbb{E}(\hat{N}_0)$	$V(\hat{N}_0)$	RRMSE	$\mathbb{E}(\hat{\tau})$	$V(\hat{\tau})$	RRMSE
50	10	10^4	88	2123,5	1,19	0,24	0,01	1,62	30,54	188,89	0,48	0,22	0,01	1,57
50	10	10^5	118	4642,2	1,93	0,24	0,01	1,62	30,58	190,32	0,48	0,22	0,01	1,57
50	10	10^6	147	7794,7	2,62	0,24	0,01	1,62	30,64	190,07	0,48	0,22	0,01	1,56
50	10	10^7	171	1.84×10^4	3,63	0,24	0,01	1,62	30,52	188,25	0,48	0,22	0,01	1,57
50	30	10^4	90	1654,2	1,14	0,31	0,01	0,27	51,54	142,57	0,24	0,32	0,01	0,30
50	30	10^5	109	3754,1	1,70	0,31	0,01	0,27	51,58	145,68	0,24	0,32	0,01	0,30
50	30	10^6	129	6864,5	2,29	0,31	0,01	0,27	51,58	145,88	0,24	0,32	0,01	0,30
50	30	10^7	144	1.16×10^4	2,86	0,31	0,01	0,27	51,56	147,25	0,24	0,32	0,01	0,30
50	50	10^4	68	1125,4	0,76	0,45	0,01	0,24	54,86	118,40	0,24	0,45	0,01	0,24
50	50	10^5	73	2127,2	1,03	0,45	0,01	0,24	54,84	119,01	0,24	0,46	0,01	0,24
50	50	10^6	78	4058,1	1,39	0,45	0,01	0,24	54,92	118,91	0,24	0,45	0,01	0,24
50	50	10^7	86	7854,0	1,91	0,45	0,01	0,24	54,92	118,91	0,24	0,45	0,01	0,24
500	10	10^4	869	1.51×10^5	1,07	0,12	0,00	0,54	514,13	4.69×10^4	0,43	0,12	0,00	0,56
500	10	10^5	1332	6.01×10^5	2,27	0,12	0,00	0,54	529,50	5.77×10^4	0,48	0,12	0,00	0,56
500	10	10^6	1748	1.27×10^6	3,36	0,12	0,00	0,54	531,44	5.81×10^4	0,49	0,12	0,00	0,56
500	10	10^7	2369	3.82×10^6	5,41	0,12	0,00	0,54	532,50	5.94×10^4	0,49	0,12	0,00	0,57
500	30	10^4	516	2808,3	0,11	0,30	0,00	0,13	501,10	1975,29	0,09	0,30	0,00	0,13
500	30	10^5	517	2964,9	0,11	0,30	0,00	0,13	500,97	1984,82	0,09	0,30	0,00	0,13
500	30	10^6	516	2841,7	0,11	0,30	0,00	0,13	500,92	1971,19	0,09	0,30	0,00	0,13
500	30	10^7	516	2839,9	0,11	0,30	0,00	0,13	501,08	1984,59	0,09	0,30	0,00	0,13
500	50	10^4	503	304,2	0,04	0,50	0,00	0,06	501,14	288,20	0,03	0,50	0,00	0,06
500	50	10^5	503	303,0	0,04	0,50	0,00	0,06	501,00	287,80	0,03	0,50	0,00	0,06
500	50	10^6	503	303,3	0,04	0,50	0,00	0,06	501,06	287,94	0,03	0,50	0,00	0,06
500	50	10^7	503	302,7	0,04	0,50	0,00	0,06	501,02	290,14	0,03	0,50	0,00	0,06

- dflat sur N_0 .
- τ basés sur la moyenne de la loi *a posteriori* ($\hat{\tau}_{moy}$) sont meilleurs avec une loi *a priori* pour $\tau \sim \beta(2,1,10)$ par rapport à une loi $\beta(3,3)$ (cf. Table 5.3). Ils sont meilleurs avec une loi Uniforme sur N_0 .
- N_0 basés sur la médiane ($\hat{N}_{0_{med}}$) sont similaires quelle que soit la loi *a priori* sur τ et sur N_0 .
- τ basés sur la médiane ($\hat{\tau}_{med}$) sont meilleurs avec une loi *a priori* pour $\tau \sim \beta(2,1,10)$ par rapport à une loi $\beta(3,3)$. Ils sont meilleurs avec une loi Uniforme sur N_0 .
- Pour $\tau > 0,1$, les estimateurs de N_0 et τ sont meilleurs avec une loi *a priori* pour $\tau \sim \beta(3,3)$ par rapport à une loi $\beta(2,1,10)$ (cf. Table 5.3).

Lorsque la distribution *a priori* de τ est une loi $\beta(3,3)$ et celle de N_0 est de la forme $1/N_0^2$:

- $\tau = 0,1$, les estimateurs sont moins bons par rapport à une loi non informative (dflat)
- $\tau > 0,1$, les estimateurs sont meilleurs par rapport à une loi non informative (dflat)

Toutefois, si on considère qu'un bon estimateur est un estimateur dont le RRMSE est inférieur ou égal à 0,2 alors on peut dire que cet objectif est atteint lorsque la loi *a priori* de τ est une loi $\beta(3,3)$ et pour N_0 une loi *a priori* dflat ou de la forme $1/N^2$ et ce, pour des taux supérieurs ou égaux à 0,3.

TABLE 5.3 : Comportement des estimateurs dans le cas convergent selon le choix de la loi *a priori* pour N_0 et τ ; τ est de loi $\beta(2,1,10)$ ou $\beta(3,3)$ et N_0 est de loi uniforme sur $[1, A]$ avec $A = \{10^5, 10^7, 1/N_0^2\}$ ou de loi une loi dflat. $\hat{N}_{0_{moy}}$ et $\hat{\tau}_{moy}$ représentent respectivement l'estimation de N_0 et de τ par la moyenne de la distribution *a posteriori*; $\hat{N}_{0_{med}}$ et $\hat{\tau}_{med}$ représentent respectivement l'estimation de N_0 et de τ par la médiane de la distribution *a posteriori*; RRMSE représente le ratio entre la racine carré de l'Erreur Quadratique Moyenne et la vraie valeur du paramètre.

N_0	τ	Prior		$\hat{N}_0 = \hat{N}_{0_{moy}}$			$\hat{\tau} = \hat{\tau}_{moy}$			$\hat{N}_0 = \hat{N}_{med}$			$\hat{\tau} = \hat{\tau}_{med}$		
		$\pi(N_0)$	$\pi(\tau)$	$\mathbb{E}(\hat{N}_0)$	$V(\hat{N}_0)$	RRMSE	$\mathbb{E}(\hat{\tau})$	$V(\hat{\tau})$	RRMSE	$\mathbb{E}(\hat{N}_0)$	$V(\hat{N}_0)$	RRMSE	$\mathbb{E}(\hat{\tau})$	$V(\hat{\tau})$	RRMSE
50	10	10^5	(2,1,10)	238,98	5786,44	4,07	0,11	0,00	0,30	64,20	413,49	0,50	0,09	0,00	0,30
50	10	10^7	(2,1,10)	298,14	1.01×10^4	5,35	0,11	0,00	0,30	64,20	412,03	0,50	0,09	0,00	0,30
50	10	10^5	(3,3)	32,08	133,15	0,43	0,30	0,01	2,19	23,13	51,32	0,56	0,29	0,01	2,18
50	10	10^7	(3,3)	32,27	135,45	0,42	0,30	0,01	2,19	23,27	50,06	0,55	0,29	0,01	2,18
50	30	10^5	(2,1,10)	271,38	9855,21	4,85	0,16	0,00	0,48	92,84	629,17	0,99	0,15	0,00	0,52
50	30	10^7	(2,1,10)	353,53	2.22×10^4	6,76	0,16	0,00	0,48	92,72	628,44	0,99	0,15	0,00	0,52
50	30	10^5	(3,3)	55,38	147,08	0,27	0,35	0,00	0,28	46,70	60,77	0,17	0,35	0,01	0,29
50	30	10^7	(3,3)	55,38	148,82	0,27	0,35	0,00	0,28	46,64	60,27	0,17	0,35	0,01	0,29
50	10	dflat	(2,1,10)	287,71	3.73×10^4	6,13	0,11	0,00	0,41	67,98	596,97	0,61	0,09	0,00	0,40
50	10	dflat	(3,3)	33,24	143,92	0,41	0,30	0,01	2,13	23,75	54,34	0,55	0,29	0,01	2,13
50	10	$1/N_0^2$	(3,3)	20,87	31,21	0,59	0,37	0,01	2,87	18,76	23,74	0,63	0,37	0,01	2,87
50	30	dflat	(2,1,10)	925,98	3.42×10^6	40,94	0,13	0,00	0,59	132,55	2299,30	1,91	0,11	0,00	0,65
50	30	dflat	(3,3)	68,31	512,03	0,58	0,30	0,01	0,26	54,17	176,20	0,28	0,30	0,01	0,28
50	30	$1/N_0^2$	(3,3)	49,97	80,42	0,18	0,36	0,00	0,30	46,10	54,17	0,17	0,36	0,01	0,31
50	50	dflat	(2,1,10)	387,78	2.24×10^6	30,70	0,29	0,01	0,45	73,92	357,77	0,61	0,30	0,01	0,45
50	50	dflat	(3,3)	57,06	106,45	0,25	0,47	0,01	0,18	52,26	28,82	0,12	0,48	0,01	0,18
50	50	$1/N_0^2$	(3,3)	52,12	24,15	0,11	0,50	0,01	0,15	50,38	17,92	0,08	0,50	0,01	0,15
500	10	dflat	(2,1,10)	3.15×10^4	6.47×10^9	172,43	0,08	0,00	0,55	2686,30	8.35×10^7	18,79	0,06	0,00	0,66
500	10	dflat	(3,3)	555,85	6.10×10^4	0,51	0,15	0,00	0,79	399,21	1.92×10^4	0,34	0,15	0,00	0,79
500	10	$1/N_0^2$	(3,3)	326,27	6659,67	0,38	0,20	0,00	1,12	293,76	4367,46	0,43	0,20	0,00	1,11
500	30	dflat	(2,1,10)	1.16×10^6	9.12×10^{12}	6471,39	0,23	0,00	0,31	590,48	1.32×10^4	0,29	0,25	0,00	0,24
500	30	dflat	(3,3)	550,02	5731,45	0,18	0,30	0,00	0,15	510,54	3863,91	0,13	0,30	0,00	0,15
500	30	$1/N_0^2$	(3,3)	505,71	3206,84	0,11	0,31	0,00	0,15	494,70	2643,25	0,10	0,31	0,00	0,15
500	50	dflat	(2,1,10)	525,98	610,27	0,07	0,48	0,00	0,07	512,26	252,94	0,04	0,48	0,00	0,07
500	50	dflat	(3,3)	511,25	230,32	0,04	0,49	0,00	0,05	503,98	211,65	0,03	0,50	0,00	0,05
500	50	$1/N_0^2$	(3,3)	504,51	210,65	0,03	0,50	0,00	0,05	502,68	203,22	0,03	0,50	0,00	0,05

5.3 Article en préparation

Stabilization of abundance estimators from removal sampling

Bioche C.¹, Bord S.², and Druilhet P.^{*1}

¹Laboratoire de mathématiques, UMR CNRS 6620 Campus des Cèzeaux, B.P. 80026, F-63171 Aubière cedex, France

²INRA, UR346 Unité Epidémiologie Animale, Centre INRA de Theix, F-63122 Saint Genès Champanelle, France

Abstract

We propose a Bayesian approach to estimate abundance from data collected by removal sampling. For unknown catchability rate, the estimation of abundance may be very unstable or divergent depending on the choice of the prior distribution. In this paper, we discuss the influence of the choice of the hyperparameter on the behaviour of posterior analysis.

1 Introduction

The removal sampling method is commonly used in Ecology to estimate the population size of animals species and the probability τ for one individual to be detected by the method of detection on each point of observation (Williams et al., 2002; Dorazio and Jelks, 2005; Dauphin et al., 2009; Brun et al., 2011). The removal sampling consist in capturing individuals on several occasions over a short period of time to preserve the assumptions of closed population. The captured individuals are removed from the population. The removal sampling data are the sequence of captures, (X_1, \dots, X_n) , observed during the n successive samplings. The main assumption of removal sampling is that the probability of capture is the same for all individuals in the population.

To estimate N and τ based on removal sampling data (X_1, \dots, X_n) , Moran (1951) was the first to propose one estimation of (N, τ) through an approximation of the maximum likelihood based on Stirling's approximation. Few years later, Zippin (1956, 1958) propose a graphical method to estimate the population size and Seber (1982) propose an estimation of confidence interval assuming asymptotic normality approximation of maximum likelihood estimate.

However, estimations based on the likelihood function failed for a lot of reasons. First, the assumption of asymptotic normality approximation of maximum likelihood estimate fail in many situations (Hirst, 1994) and computed confidence interval could be unsatisfactory as the lower limit could be less than the total number of capture observed (Otis et al., 1978; Borchers et al., 2002).

*Pierre.DRUILHET@univ-bpclermont.fr

Hence, Hirst (1994) proposed confidence interval based on the profile likelihood ratio and Carle and Strub (1978) weighted the likelihood by beta distribution and consider the maximum of marginal likelihood to estimate the population size. Moreover, the likelihood function conduct to infinite estimations of population size N with a non null probability (Carle and Strub, 1978; Bolfarine et al., 1992; Bedrick, 1994). Bedrick (1994) proposed a criterion based on the number of consecutive sampling realised and the cumulative number of individuals detected to determinate if the estimation of N was finite *i.e.* if the maximum of likelihood function converge. This results was previously conjectured by Carle and Strub (1978). Furthermore, when methods based on the likelihood function success to converge, the maximum likelihood estimate will favor a small N (Schnute, 1983). In this context, Bayesian methods seems to be alternatives methods to estimate both N and p Bolfarine et al. (1992); Schwarz and Seber (1999). One advantage of Bayesian approaches is to take into account available informations about the distribution of parameters to estimate through the prior distributions.

However, in the absence of knowledge or information about the distribution of parameters N and p , it is the most intuitive to use the uniform distribution on $[0, 1]$ on p as a non-informative or a flat prior ; it is a non-informative that all possible values in $[0, 1]$ have the same probability so they are equally likely *a priori* (Robert, 2005). Beta(a, b) distribution could be also used with $a = b = 0$ corresponds to a uniform distribution for $\text{logit}(p)$ (Haldane, 1932), or $a = b = 1/2$ Jeffrey's prior which corresponds to a uniform distribution for $\sin^{-1} \sqrt{p}$ (Jeffreys, 1946).

However, flat prior diverge. With vague prior on hyperparameters, the estimates based on the mean or on the median of posterior distribution are sensitive to the choice of prior distribution.

The objective of the present study is to give recommendations about the choice of prior for N and p according to the level of p .

2 Removal sampling likelihood and Poisson iid model proximity

We consider a population with N_0 individuals. We do many captures and, at each capture, we caught a number x_i of animals. We assume that the probability of capture, τ , is constant across individuals and trappings and that individuals are captured independently. We consider a binomial model, *i.e.* we assume that x_1 is an observation of the variable X_1 where $X_1 \sim \mathcal{B}(N_0, \tau)$. We also assume that $X_2|X_1 \sim \mathcal{B}(N_0 - X_1, \tau)$, $X_3|(X_1, X_2) \sim \mathcal{B}(N_0 - X_1 - X_2, \tau)$, etc. After k captures, we have the vector of observations $x = (x_1, x_2, \dots, x_k)$, and the likelihood is

$$\begin{aligned} L((N_0, \tau); x) &= \prod_{i=1}^k P(X_i | x_1, \dots, x_{i-1}) \\ &= \frac{N_0!}{(N_0 - T)! \prod_{i=1}^k x_i!} \tau^T (1 - \tau)^{k(N_0 - T) + d_0} \end{aligned} \quad (1)$$

where $T = \sum_{i=1}^k x_i$ and $d_0 = \sum_{i=1}^k (i-1)x_i$.

Note that only the part $(1-\tau)^{d_0}$ depends on the order of the captures. When τ is near 0, $(1-\tau)^{d_0}$ is near 1 so is negligible.

In this part, we state that the likelihood of the removal sampling model tends to the likelihood of a Poisson iid model. Intuitively, when τ is near 0, i.e. the probability of capture is very small, and N_0 is very large, we catch few animals at each capture so $N_0 \approx N_0 - \sum_{i=1}^{k-1} x_i$. So, X_1, X_2, \dots, X_k behave similarly to k independant random variables with Binomial distribution $\mathcal{B}(N_0, \tau)$. Moreover, when the number of trials N_0 is large, and when the probability of capture τ is close to zero, the binomial distribution $\mathcal{B}(N_0, \tau)$ can be approximated by a Poisson distribution with parameter $N_0\tau$. Therefore we have:

Proposition 2.1. *Consider that, after k captures, we have the vector of observations $x = (x_1, x_2, \dots, x_k)$. Assume that τ goes to 0, N_0 goes to $+\infty$ and there exists $\lambda > 0$ such that τN_0 goes to λ . Then,*

$$L((N_0, \tau); x) \xrightarrow[N_0\tau \rightarrow \lambda]{N_0 \rightarrow +\infty} L_p(\lambda; x)$$

where $L_p(\lambda; x) = \prod_{i=1}^k e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}$ is the likelihood of k independant Poisson distributions with parameter λ .

Proof. We have $L((N_0, \tau); x) = \frac{N_0! \tau^T (1-\tau)^{k(N_0-T)+d_0}}{(N_0-T)! \prod_{i=1}^k x_i!}$. We make the change of variable $\tau = \frac{\lambda}{N_0}$, and we study what happen when N_0 goes to $+\infty$. We have

$$L\left((N_0, \frac{\lambda}{N_0}); x\right) = \frac{N_0!}{(N_0-T)! N_0^T} \times \left(1 - \frac{\lambda}{N_0}\right)^{-kT+d_0} \quad (2)$$

$$\times \frac{\lambda^T}{\prod x_i!} \times \left(1 - \frac{\lambda}{N_0}\right)^{kN_0} \quad (3)$$

where the two right-hand side terms of the ligne (2) go to 1 when $N_0 \rightarrow +\infty$.

And in the line (3), $\left(1 - \frac{\lambda}{N_0}\right)^{kN_0} \xrightarrow[N_0 \rightarrow +\infty]{} e^{-k\lambda}$. So, $L\left((N_0, \frac{\lambda}{N_0}); x\right) \xrightarrow[N_0 \rightarrow +\infty]{} e^{-k\lambda} \frac{\lambda^T}{\prod x_i!} = L_p(\lambda; x)$. \square

Usually, when the parameters go to the extremities of their interval of definition, the likelihood goes to 0. Here, it is not the case, the likelihood converges to the likelihood of an other model. Note that the Poisson iid limiting model, is not identifiable. This feature will explain some behaviour of both the frequentist and bayesian analysis.

3 Frequentist analysis of removal sampling

Now, we consider a frequentist approach to estimate N_0 and τ . Moran (1951) was the first to propose a multinomial model in view to estimate simultaneously

N_0 and τ through the asymptotic properties of the MLE. A maximum likelihood approach often leads to infinite estimations for N_0 (Carle and Strub, 1978; Bolfarine et al., 1992; Schnute, 1983).

Note that $\max_{(N_0, \tau)} L((N_0, \tau); x) = \max_{N_0} (\max_{\tau} L((N_0, \tau); x))$. For each given N_0 , the likelihood $L((N_0, \tau); x)$ is maximized $\hat{\tau}(N_0) = T\{k(N_0 - T) + T + d_0\}^{-1}$. Replacing τ in Formula (1) by $\hat{\tau}(N_0)$, the MLE satisfies $\hat{\tau} = \hat{\tau}(\hat{N}_0)$, where \hat{N}_0 maximizes the profile likelihood $L((N_0, \hat{\tau}(N_0)); x)$. Bedrick (1994) gives a necessary and sufficient condition for the existence of the MLE estimator:

Proposition 3.1 (Bedrick (1994)).

- The profile log-likelihood $\log L((N_0, \hat{\tau}(N_0)); x)$ is an increasing function for $N_0 \geq T$ when $d_0 \geq 0.5T(k-1) + 0.5k$, and is a decreasing function of N_0 when $d_0 \leq T[(1+2T)^{1/k} - 1]^{-1}$.
- The profile log-likelihood $\log L((N_0, \hat{\tau}(N_0)); x)$ is maximized at a finite value if and only if $d_0 < 0.5T(k-1) + 0.5k$.
- The derivative of $\log L((N_0, \hat{\tau}(N_0)); x)$ has at most one finite root for $N_0 > T$.
- The profile log-likelihood $\log L((N_0, \hat{\tau}(N_0)); x)$ is unimodal, and provides a unique MLE for N_0 .

The convergence of the likelihood of the removal sampling model to the likelihood of the Poisson iid model (Proposition 2.1) still holds for the profile likelihood.

Proposition 3.2. The profile likelihood satisfies the following convergence:

$$L((N_0, \hat{\tau}(N_0)); x) \xrightarrow{N_0 \rightarrow +\infty} L_p(\hat{\lambda}; x)$$

where $\hat{\lambda} = \bar{x} = \frac{1}{k} \sum_{i=1}^k x_i$ is the ML estimator in the Poisson iid model.

Proof. We have

$$L((N_0, \hat{\tau}(N_0)); x) = \frac{N_0! k^T}{(N_0 - T)! (k(N_0 - T) + T + d_0)^T} \quad (4)$$

$$\times \left(1 - \frac{T}{k(N_0 - T) + T + d_0}\right)^{-kT + d_0} \quad (5)$$

$$\times \frac{1}{\prod x_i!} \left(\frac{T}{k}\right)^T \times \left(1 - \frac{T}{k(N_0 - T) + T + d_0}\right)^{kN_0} \quad (6)$$

where the right-hand side of line (4) and line (5) go to 1 when N_0 goes to $+\infty$.

And, in line (6), $\left(1 - \frac{T}{k(N_0 - T) + T + d_0}\right)^{kN_0} \xrightarrow{N_0 \rightarrow \infty} e^{-T}$. So $L((N_0, \hat{\tau}(N_0)); x) \xrightarrow{N_0 \rightarrow \infty} \frac{1}{\prod x_i!} \left(\frac{T}{k}\right)^T e^{-k\frac{T}{k}} = L_p(\hat{\lambda}; x)$ with $\hat{\lambda} = \frac{T}{k}$. \square

Corollary 3.1. Assume that $\hat{N}_{0,MV} = +\infty$. Then, the maximum likelihood value is $L_p(\hat{\lambda}_{MV}; x)$ where $\hat{\lambda}_{MV} = \frac{T}{k}$. Moreover, there exists a sequence $\{(N_{0,n}, \tau_n)\}$ such that $L((N_{0,n}, \tau_n); x) \xrightarrow{n \rightarrow +\infty} L_p(\hat{\lambda}_{MV}; x)$ and $N_{0,n} \tau_n \xrightarrow{n \rightarrow +\infty} \hat{\lambda}_{MV}$.

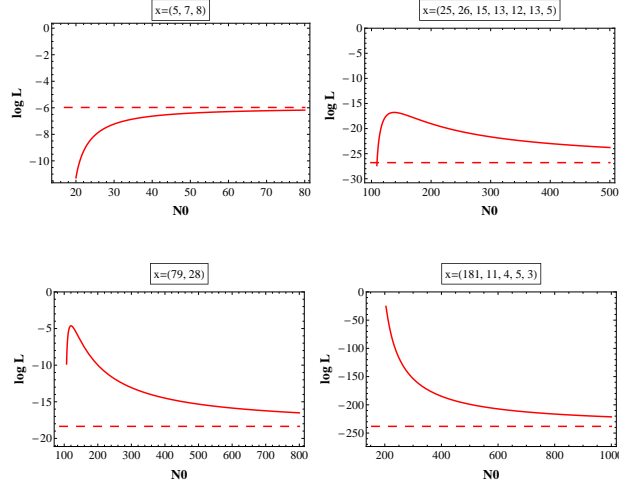


Figure 1: Four examples of profile log-likelihoods in continuous line and their limit profile log-likelihood in dashed line.

Proof. If $\hat{N}_{0,MV} = +\infty$, from Proposition 3.2, the maximum likelihood value is $L_p(\hat{\lambda}_{MV}; x)$ where $\hat{\lambda}_{MV} = \bar{x}$. By definition, $(\hat{N}_{0,MV}, \hat{\tau}_{MV}) = \text{ArgMax}_{(N_0, \tau)} L((N_0, \tau); x)$. So, there exists a sequence $\{(N_{0,n}, \tau_n)\}$ which goes to $(\hat{N}_{0,MV}, \hat{\tau}_{MV})$ and such that $L((N_{0,n}, \tau_n); x) \xrightarrow{n \rightarrow +\infty} L_p(\hat{\lambda}_{MV}; x)$. From Proposition 2.1, there exists N such that for $n > N$, $L((N_{0,n}, \tau_n); x)$ can be approximated by $L_p(N_{0,n}\tau_n; x)$. Thus, $L_p(N_{0,n}\tau_n; x) \xrightarrow{n \rightarrow +\infty} L_p(\hat{\lambda}_{MV}; x)$. By continuity of L_p , $\lim_{n \rightarrow +\infty} L_p(N_{0,n}\tau_n; x) = L_p(\lim_{n \rightarrow +\infty} N_{0,n}\tau_n; x)$. But $\lim_{n \rightarrow +\infty} L_p(N_{0,n}\tau_n; x) = L_p(\hat{\lambda}_{MV}; x)$. And from unicity of the ArgMax, $\lim_{n \rightarrow +\infty} N_{0,n}\tau_n = \hat{\lambda}_{MV}$. \square

In Figure 1, we consider four examples as in Bedrick (1994) with the addition of the asymptotic behaviour. The first is composed of three trappings of mottlet sculpin (Carle and Strub, 1978). The second deals with seven trappings of whitefish (Otis et al., 1978). The third case gives results of two trappings of trout (Seber and Le Cren, 1967). And the last one is composed of five trappings of mayflies (Otis et al., 1978).

4 Bayesian analysis of removal sampling

We now consider the problem with a Bayesian approach. We put a prior Π on (N_0, τ) . We denote by π the density function (w.r.t. the product measure

$\mu = \left(\sum_{i \in \mathbb{N}^*} \delta_i \right) \times \lambda_{[0,1]}$. If Π is a probability measure, we can use the Bayes Formula to write the posterior density function:

$$\pi(N_0, \tau | x) = \frac{f(x | N_0, \tau) \pi(N_0, \tau)}{\int \int f(x | N_0, \tau) \pi(N_0, \tau) d\mu(N_0, \tau)}. \quad (7)$$

It is usual in Bayesian litterature to write:

$$\pi(N_0, \tau | x) \propto f(x | N_0, \tau) \pi(N_0, \tau). \quad (8)$$

If Π is an improper measure but $\int \int f(x | N_0, \tau) \pi(N_0, \tau) d\mu(N_0, \tau) < +\infty$, we can formally applied Formula (7) to get a posterior probability. Sometimes, authors use improper priors such as a flat prior on N_0 . In this section, we will consider different priors, proper or improper, which appear in the litterature and we will study the posterior analysis. We will see that improper priors can give proper or improper posteriors. When the posterior is proper, we can study the posterior estimator of N_0 . We also give the case of a proper prior which give a proper posterior but an infinite estimator.

4.1 Convergence of estimators of N_0

To begin, because two parameterizations are interesting for our problem, we establish propositions about change of parameterization. In order to use the result of Section 2, we consider the parameterization (N_0, λ) where $(N_0, \lambda) = (N_0, N_0 \tau)$. If Π is a prior on (N_0, τ) , the corresponding prior on (N_0, λ) is denoted by $\tilde{\Pi}$.

Lemma 4.1. *The prior density is $\tilde{\pi}(N_0, \lambda) = \frac{1}{N_0} \pi\left(N_0, \frac{\lambda}{N_0}\right) \mathbb{1}_{[0, N_0]}(\lambda)$ and similarly $\tilde{\pi}(N_0, \lambda | x) = \frac{1}{N_0} \pi\left(N_0, \frac{\lambda}{N_0} | x\right) \mathbb{1}_{[0, N_0]}(\lambda)$.*

Proof. We have $\tilde{\Pi}(N_0, [a, b]) = \Pi(N_0, [\frac{a}{N_0}, \frac{b}{N_0}]) = \int_{\frac{a}{N_0}}^{\frac{b}{N_0}} \pi(N_0, \tau) d\tau = \int_a^b \pi(N_0, \frac{\lambda}{N_0}) \frac{1}{N_0} d\lambda$ by the change of variable $\tau = \frac{\lambda}{N_0}$. The result follows. Similarly, we can obtain $\tilde{\pi}(N_0, \lambda | x) = \frac{1}{N_0} \pi\left(N_0, \frac{\lambda}{N_0} | x\right) \mathbb{1}_{[0, N_0]}(\lambda)$. \square

The marginal posterior distribution of N_0 can be defined either from Π by $\pi(N_0 | x) = \int_0^1 \pi(N_0, \tau | x) d\tau$ or equivalently from $\tilde{\Pi}$ by $\pi(N_0 | x) = \int_0^{N_0} \tilde{\pi}(N_0, \lambda | x) d\lambda$. We now give a technical lemma which will be usefull latter in the article.

Lemma 4.2. *For $a > 0$ and $b > 0$, we have*

$$N_0^a \int_0^1 \tau^{a-1} (1-\tau)^{b-1} L((N_0, \tau); x) d\tau \xrightarrow{N_0 \rightarrow +\infty} K_{a,T} > 0$$

with $K_{a,T} = \int_0^{+\infty} \lambda^{a-1} L_p(\lambda; x) d\lambda = \frac{1}{\pi x i!} \frac{\Gamma(T+a)}{k^{T+a}}$.

Proof. By the change of variable $\tau = \lambda N_0^{-1}$, we have $N_0^a \int_0^1 \tau^{a-1} (1-\tau)^{b-1} L((N_0, \tau); x) d\tau = N_0^a \int_0^{N_0} \left(\frac{\lambda}{N_0}\right)^{a-1} \left(1 - \frac{\lambda}{N_0}\right)^{b-1} L((N_0, \frac{\lambda}{N_0}); x) \frac{1}{N_0} d\lambda = \int_0^{N_0} \lambda^{a-1} \left(1 - \frac{\lambda}{N_0}\right)^{b-1} L\left((N_0, \frac{\lambda}{N_0}); x\right) d\lambda$. And, for any λ , $\lambda^{a-1} \left(1 - \frac{\lambda}{N_0}\right)^{b-1}$

$L\left((N_0, \frac{\lambda}{N_0}); x\right) \mathbb{1}_{[0, N_0]}(\lambda) \leq \lambda^{T+a-1} e^{-k\lambda}$ which is an integrable function. Moreover, from Proposition 2.1, $L\left((N_0, \frac{\lambda}{N_0}); x\right) \mathbb{1}_{[0, N_0]}(\lambda)$ converges to $L_p(\lambda; x)$ when N_0 goes to $+\infty$. So, from dominated convergence theorem,

$$\int_0^{N_0} \lambda^{a-1} \left(1 - \frac{\lambda}{N_0}\right)^{b-1} L\left((N_0, \frac{\lambda}{N_0}); x\right) d\lambda \xrightarrow{N_0 \rightarrow +\infty} \int_0^{+\infty} \lambda^{a-1} L_p(\lambda; x) d\lambda.$$

We have $\int_0^{+\infty} \lambda^{a-1} L_p(\lambda; x) d\lambda = \frac{1}{\pi x!} \int_0^{+\infty} \lambda^{T+a-1} e^{-k\lambda} d\lambda$. But we know that $\int_0^{+\infty} \lambda^{T+a-1} e^{-k\lambda} d\lambda = \Gamma(T+a) k^{-(T+a)}$ because we recognize the density of a Gamma distribution $\Gamma(T+a, k)$. The result follows. \square

The following theorem gives necessary and sufficient conditions on the parameters of a prior distribution of the form $\pi(N_0, \tau) \propto \frac{1}{N_0^c} \times \tau^{a-1} (1-\tau)^{b-1}$ to have a proper posterior and a well defined Bayes estimator on N_0 .

Theorem 4.1. *Consider a prior Π on (N_0, τ) whose density satisfies $\pi(N_0, \tau) \propto \frac{1}{N_0^c} \times \tau^{a-1} (1-\tau)^{b-1}$. Then,*

1. *the posterior distribution $\Pi(N_0, \tau|x)$ is proper if and only if $a+c > 1$,*
2. *the Bayes estimator of N_0 , $\mathbb{E}_\Pi(N_0|x)$, is defined if and only if $a+c > 2$.*

Proof.

1. From Formula 8, $\pi(N_0, \tau|x) \propto L((N_0, \tau); x) \pi(N_0, \tau)$, i.e. $\pi(N_0, \tau|x) \propto L((N_0, \tau); x) \frac{1}{N_0^c} \tau^{a-1} (1-\tau)^{b-1}$. So, $\int \int \pi(N_0, \tau|x) d\tau dN_0 \propto \int \frac{1}{N_0^c} \int L((N_0, \tau); x) \tau^{a-1} (1-\tau)^{b-1} d\tau dN_0$. From Lemma 4.2, $N_0^a \int L((N_0, \tau); x) \tau^{a-1} (1-\tau)^{b-1} d\tau \rightarrow K_{a,T}$. So, for $\varepsilon > 0$, there exists N such that for $N_0 > N$, $\frac{(K_{a,T}-\varepsilon)}{N_0^a} \leq \int_0^1 L((N_0, \tau); x) \tau^{a-1} (1-\tau)^{b-1} d\tau \leq \frac{(K_{a,T}+\varepsilon)}{N_0^a}$, i.e. $\frac{(K_{a,T}-\varepsilon)}{N_0^{a+c}} \leq \frac{1}{N_0^c} \int_0^1 L((N_0, \tau); x) \tau^{a-1} (1-\tau)^{b-1} d\tau \leq \frac{(K_{a,T}+\varepsilon)}{N_0^{a+c}}$. So, $\int_0^{+\infty} \frac{(K_{a,T}-\varepsilon)}{N_0^{a+c}} dN_0 \leq \int_0^{+\infty} \frac{1}{N_0^c} \int_0^1 L((N_0, \tau); x) \tau^{a-1} (1-\tau)^{b-1} d\tau dN_0 \leq \int_0^{+\infty} \frac{(K_{a,T}+\varepsilon)}{N_0^{a+c}} dN_0$, i.e. $\int_0^{+\infty} \frac{(K_{a,T}-\varepsilon)}{N_0^{a+c}} dN_0 \leq \int_0^{+\infty} \frac{(K_{a,T}+\varepsilon)}{N_0^{a+c}} dN_0$. Thus, $\Pi(N_0, \tau|x)$ is proper if and only if $a+c > 1$.
2. The Bayes estimator of N_0 is defined by $\mathbb{E}_\Pi(N_0|x) = \int N_0 \pi(N_0|x) dN_0 = \int N_0 \int \pi(N_0, \tau|x) d\tau dN_0 \propto \int N_0 \int L((N_0, \tau); x) \pi(N_0, \tau) d\tau dN_0 \propto \int \frac{1}{N_0^{c-1}} \int L((N_0, \tau); x) \tau^{a-1} (1-\tau)^{b-1} d\tau dN_0$. Similarly to the point 1., we obtain, $\int_0^{+\infty} \frac{(K_{a,T}-\varepsilon)}{N_0^{a+c-1}} dN_0 \leq \mathbb{E}_\Pi(N_0|x) \leq \int_0^{+\infty} \frac{(K_{a,T}+\varepsilon)}{N_0^{a+c-1}} dN_0$. Thus $\mathbb{E}_\Pi(N_0|x) < +\infty$ if and only if $a+c-1 > 1$, i.e. $a+c > 2$.

\square

We now review different usual priors, proper or improper, on (N_0, τ) . For example we consider a flat prior on (N_0, τ) or a flat prior on N_0 with priors on τ which encourage τ to be close to 0 as Jeffreys or Haldane prior. We also consider priors which penalize large values of N_0 . First we search if the posteriors are defined or not and, when the posterior distribution is defined, we look at the posterior estimator of N_0 .

Corollary 4.1.

1. If $\pi(N_0, \tau) \propto 1$, then the posterior $\Pi(N_0, \tau|x)$ is an improper measure.
2. If $\pi(N_0, \tau) \propto \frac{1}{N_0}$, then
 - (a) the posterior $\Pi(N_0, \tau|x)$ is a proper measure,
 - (b) the Bayes estimator of N_0 satisfies $\mathbb{E}(N_0|x) = +\infty$.
3. If $\pi(N_0, \tau) \propto \frac{1}{N_0^2}$, then
 - (a) the posterior $\Pi(N_0, \tau|x)$ is a proper measure,
 - (b) the Bayes estimator of N_0 , $\mathbb{E}(N_0|x)$, converges.
4. If $\pi(N_0, \tau) \propto \frac{1}{\tau(1-\tau)}$, then the posterior $\Pi(N_0, \tau|x)$ is an improper measure.
5. If $\pi(N_0, \tau) \propto \frac{1}{\sqrt{\tau(1-\tau)}}$, then the posterior $\Pi(N_0, \tau|x)$ is an improper measure.
6. If $\pi(N_0, \tau) \propto \frac{1}{N_0^{\frac{3}{2}}} \times \frac{1}{\sqrt{\tau(1-\tau)}}$, then
 - (a) the posterior $\Pi(N_0, \tau|x)$ is a proper measure,
 - (b) the posterior estimator satisfies $\mathbb{E}(N_0|x) = +\infty$.

Proof. All these results are direct consequences of Theorem 4.1 with different values for a , b and c . \square

We note that improper priors often lead to improper posteriors or to infinite estimators. So, these priors don't seem to be adapted to the situation. We may think that consider proper priors which "approach" them could be a solution. That is the subject of Section 4.2.

4.2 Analysis with vague proper priors

We now consider the case of sequences of proper priors which approximate improper priors. To begin, we explain how we are using the term "approximate". For example, we say that a sequence of probabilities Π_n defined by $\pi_n(N_0, \tau) \propto \mathbb{1}_{\{1 \leq N_0 \leq n\}}$ approximate an improper measure Π on (N_0, τ) such that $\pi(N_0, \tau) \propto 1$. More generally, we use the definition of q -vague convergence given by Bioche and Druilhet (2014): the sequence $\{\Pi_n\}_{n \in \mathbb{N}}$ converges q -vaguely to Π or approximate Π if there exists a sequence of positive real numbers $\{a_n\}_{n \in \mathbb{N}}$ such that $\{a_n \Pi_n\}_{n \in \mathbb{N}}$ converges vaguely to Π . In this article, we define a sequence of vague priors as a sequence of probability measures that converges q -vaguely to an improper measure. First, we state that the q -vague convergence of the priors on N_0 entails the q -vague convergence of the marginal posteriors on N_0 .

Proposition 4.1. Assume that $\Pi_n(N_0, \tau) = \Pi_n^{(1)}(N_0) \times \Pi_n^{(2)}(\tau)$ where $\{\Pi_n^{(1)}\}$ is a sequence of vague priors which converges to the improper prior $\Pi^{(1)}$. Then, the sequence of marginal posterior distributions on N_0 , $\{\Pi_n(\cdot|x)\}$, converges q -vaguely to $\Pi(\cdot|x)$.

Proof. By definition, $\{\Pi_n(\cdot|x)\}$ converges q -vaguely to $\Pi(\cdot|x)$ if and only if there exists a sequence of positive real numbers $\{a_n\}$ such that $\{a_n \Pi_n(\cdot|x)\}$ converges vaguely to $\Pi(\cdot|x)$. Let h be a continuous function on \mathbb{N} with a compact support. We have $\int h(N_0) d\Pi_n(N_0|x) = \int h(N_0) \pi_n(N_0|x) dN_0 = \int h(N_0) \left(\int_0^1 L((N_0, \tau); x) \pi^{(2)}(\tau) d\tau \right) \pi_n^{(1)}(N_0) dN_0$. And, $N_0 \mapsto \int_0^1 L((N_0, \tau); x) \pi^{(2)}(\tau) d\tau$ is a continuous function because $\int_0^1 L((N_0, \tau); x) \pi^{(2)}(\tau) d\tau < +\infty$ for all N_0 . So, $N_0 \mapsto h(N_0) \left(\int_0^1 L((N_0, \tau); x) \pi^{(2)}(\tau) d\tau \right)$ is a continuous function with compact support. But, $\{\Pi_n^{(1)}\}$ converges q -vaguely to $\Pi^{(1)}$. So, there exists a sequence of positive real numbers $\{a_n\}$ such that $a_n \int h(N_0) \int_0^1 L((N_0, \tau); x) \pi^{(2)}(\tau) d\tau \pi_n^{(1)}(N_0) dN_0$ converges to $\int h(N_0) \int_0^1 L((N_0, \tau); x) \pi^{(2)}(\tau) d\tau \pi^{(1)}(N_0) dN_0$, i.e. $a_n \int h(N_0) d\Pi_n(N_0|x) \rightarrow \int h(N_0) d\Pi(N_0|x)$. The result follows. \square

From Proposition 4.1, if we approximate an improper prior on N_0 by a sequence of vague priors, we obtain the q -vague convergence of the sequence of the marginal posteriors on N_0 . We now state that if the limiting marginal posterior on N_0 is improper, the limiting Bayes estimator of N_0 diverges.

Proposition 4.2. *Let $\{\Pi_n(\cdot|x)\}_{n \in \mathbb{N}}$ be a sequence of probability measures on $N_0 \in \mathbb{N}$ which converges q -vaguely to an improper prior $\Pi(\cdot|x)$. Then, $\mathbb{E}_{\Pi_n}(N_0|x) \xrightarrow{n \rightarrow +\infty} +\infty$.*

Proof. From Proposition 2.8 by Bioche and Druilhet (2014), when a sequence of vague priors is used to approximate an improper prior, the mass tends to concentrate outside any compact set. For a discrete parameter, a compact set is a finite set. For all $A > 0$, we denote by C_A the set $\{0, \dots, A\}$, and we have $\Pi_n(C_A^c|x) \rightarrow 1$. So, for any $A > 0$, there exists n_A^* such that for $n > n_A^*$, $\Pi_n(C_A^c|x) \geq \frac{1}{2}$. We have $\mathbb{E}_{\Pi_n}(N_0|x) \geq \int_{N_0 > A} N_0 \Pi_n(N_0|x) \geq A \Pi_n(C_A^c|x)$. So, for $n > n_A^*$, $\mathbb{E}_{\Pi_n}(N_0|x) \geq \frac{A}{2}$. The result follows. \square

And we can establish a similar proposition about the median of the posterior distribution of N_0 :

Proposition 4.3. *Let $\{\Pi_n(\cdot|x)\}_{n \in \mathbb{N}}$ be a sequence of probability measures on $N_0 \in \mathbb{N}$ which converges q -vaguely to an improper prior $\Pi(\cdot|x)$. Then, $\text{med}_{\Pi_n}(N_0|x) \xrightarrow{n \rightarrow +\infty} +\infty$.*

Proof. Similarly to the proof of Proposition 4.2, we denote by C_A the set $\{0, \dots, A\}$ and we can state that for any $A > 0$, there exists n_A^* such that for $n > n_A^*$, $\Pi_n(C_A^c|x) \geq a > \frac{1}{2}$. Then, for $n > n_A^*$, $\text{med}_{\Pi_n}(N_0|x) \geq A$. The result follows. \square

We now consider the particular case where $\pi(N_0, \tau) \propto \frac{1}{N_0^c} \times \tau^{a-1} (1-\tau)^{b-1}$. and the sequence of vague priors is $\pi_n(N_0, \tau) \propto \frac{1}{N_0^c} \mathbb{1}_{N_0 < n} \times \tau^{a-1} (1-\tau)^{b-1}$. The following proposition gives the conditions of convergence of the Bayes estimator of N_0 .

Proposition 4.4. *Assume that $\pi(N_0, \tau) \propto \frac{1}{N_0^c} \times \tau^{a-1} (1-\tau)^{b-1}$. Put $\pi_n(N_0, \tau) \propto \frac{1}{N_0^c} \mathbb{1}_{N_0 < n} \times \tau^{a-1} (1-\tau)^{b-1}$. Then, $\mathbb{E}_{\Pi_n}(N_0|x) < +\infty$ if and only if $a + c > 2$.*

Proof. We have $\mathbb{E}_{\Pi_n}(N_0|x) = \int_0^{+\infty} N_0 \Pi(N_0|x) dN_0 \propto \int_0^n \frac{1}{N_0^{a-1}} \int_0^1 L((N_0, \tau); x) \tau^{a-1} (1-\tau)^{b-1} d\tau dN_0$. But, from Lemma 4.2, for $\varepsilon > 0$, there exists N such that for $N_0 > N$, $\frac{(K_{a,x}-\varepsilon)}{N_0^a} \leq \int_0^1 L((N_0, \tau); x) \tau^{a-1} (1-\tau)^{b-1} d\tau \leq \frac{(K_{a,x}+\varepsilon)}{N_0^a}$. So, for large values of n , $\int_N^n \frac{(K_{a,x}-\varepsilon)}{N_0^{a+c-1}} dN_0 \leq \mathbb{E}_{\Pi_n}(N_0|x) \leq \int_N^n \frac{(K_{a,x}+\varepsilon)}{N_0^{a+c-1}} dN_0$. The result follows by taking the limit when n goes to $+\infty$. \square

To conclude, we have seen that using vague priors to approximate improper priors doesn't solve all problems. Indeed, estimators are very sensitive to the choice of hyperparameters of vague priors.

References

- Bedrick, E. J. (1994). Maximum-Likelihood Estimation for the Removal Method. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 22(2):285–293.
- Bioche, C. and Druilhet, P. (2014). Approximation of improper prior by vague priors. *Submitted*.
- Bolfarine, H., Leite, J. G., and Rodrigues, J. (1992). On the Estimation of the Size of a Finite and Closed Population. *Biometrical Journal*, 34(5):577–593.
- Borchers, D. L., Buckland, S. T., and Zucchini, W. (2002). *Estimating Animal Abundance: Closed Populations*. Statistics for Biology and Health. Springer, London.
- Brun, M., Abraham, C., Jarry, M., Dumas, J., Lange, F., and Prevost, E. (2011). Estimating an homogeneous series of a population abundance indicator despite changes in data collection procedure: A hierarchical Bayesian modelling approach. *Ecological Modelling*, 222(5):1069–1079.
- Carle, F. L. and Strub, M. R. (1978). A New Method for Estimating Population Size from Removal Data. *Biometrics*, 34(4):621–630.
- Dauphin, G., Prevost, E., Adams, C. E., and Boylan, P. (2009). A Bayesian approach to estimating Atlantic salmon fry densities using a rapid sampling technique. *Fisheries Management and Ecology*, 16(5):399–408.
- Dorazio, R. M. and Jelks, H. L. (2005). Improving removal-based estimates of abundance by sampling a population of spatially distinct subpopulations. *Biometrics*, 61(4):1093–1101.
- Haldane, J. B. S. (1932). A note on inverse probability. *Mathematical Proceedings of the Cambridge Philosophical Society*, 28(01):55–61.
- Hirst, D. (1994). An Improved Removal Method for Estimating Animal Abundance. *Biometrics*, 50(2):501–505.
- Jeffreys, H. (1946). An Invariant Form for the Prior Probability in Estimation Problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 186(1007):453–461.

- Moran, P. A. P. (1951). A Mathematical Theory of Animal Trapping. *Biometrika*, 38(3/4):307–311.
- Otis, D. L., Burnham, K. P., White, G. C., and Anderson, D. R. (1978). *Statistical inference from capture data on closed animal populations*, volume 62. wildlife society.
- Robert, C. P. (2005). *Le choix bayésien: Principes et pratique*. Statistique et probabilités appliquées. Springer.
- Schnute, J. (1983). A NEW APPROACH TO ESTIMATING POPULATIONS BY THE REMOVAL METHOD. *Canadian Journal of Fisheries and Aquatic Sciences*, 40(12):2153–2169.
- Schwarz, C. J. and Seber, G. A. F. (1999). Estimating animal abundance: Review III. *Statistical Science*, 14(4):427–456.
- Seber, G. and Le Cren, E. (1967). Estimating population parameters from catches large relative to the population. *J. Animals Ecol.*, 36:631–643.
- Seber, G. A. F. (1982). The estimation of animal abundance: Griffin.
- Williams, B. K., Nichols, J. D., and Conroy, M. J. (2002). *Analysis and Management of Animal Populations*. Academic Press, San Diego, USA & London, UK, 1st edition.
- Zippin, C. (1956). An Evaluation of the Removal Method of Estimating Animal Populations. *Biometrics*, 12(2):163–189.
- Zippin, C. (1958). The Removal Method of Population Estimation. *The Journal of Wildlife Management*, 22(1):pp. 82–90.

Discussion générale et perspectives

L'estimation fiable de l'abondance d'une espèce est un outil majeur en écologie et en épidémiologie pour l'étude de la dynamique des populations d'espèces animales et la mise en évidence des facteurs explicatifs des niveaux d'abondance. Cette estimation est d'autant plus importante qu'elle concerne des espèces impliquées dans la transmission de multiples agents pathogènes comme les tiques *I. ricinus* transmettant les bactéries *Borrelia burgdorferi sensu lato*. Cependant, disposer d'une estimation fiable de l'abondance peut s'avérer compliqué car, le plus souvent, l'outil d'observation ne permet pas une détection parfaite des individus présents dans l'unité d'observation (taux de détection inférieur à 1) ; de plus, le niveau de détection est rarement constant selon les conditions d'échantillonnage (MacKenzie et al., 2002; Royle, 2006).

L'objectif global de la thèse était d'estimer par une approche bayésienne l'abondance des tiques à partir de données de « removal sampling » en présence de variations du taux d'échantillonnage.

D'un point de vue de l'écologie et de l'épidémiologie, (i) nous avons montré que pour étudier la variabilité du taux d'échantillonnage, la méthode d'échantillonnage par « removal sampling » était adaptée aux tiques contrairement aux méthodes par Capture Marquage Recapture pour lesquelles une mise au point des techniques est nécessaire, notamment pour limiter ou caractériser l'effet du marquage sur le comportement des tiques marquées ; (ii) nous avons proposé une modélisation hiérarchique bayésienne pour estimer simultanément la taille de la population de chaque unité d'observation et le taux d'échantillonnage des tiques par la méthode du drapeau en fonction des conditions d'échantillonnage. D'un point de vue théorique, cette étude nous a permis d'étudier la stabilité des estimateurs bayésiens dans le cadre de données de « removal sampling » et ceci en l'absence de connaissance des distributions *a priori* sur les paramètres.

Dans la section 6.1, nous discuterons de la pertinence de l'approche par « removal sampling » appliquée aux tiques *I. ricinus* pour l'estimation de l'abondance et de la variabilité du taux d'échantillonnage. Nous aborderons ensuite, dans la section 6.2, des valeurs estimées du taux d'échantillonnage par rapport aux estimations présentées dans la littérature. Dans la section 6.3, nous aborderons la question de la prise en

compte des variations du taux d'échantillonnage pour l'estimation de l'abondance. Nous discuterons des choix délicats des lois *a priori* dans la section 6.4. Enfin, nous traiterons, dans la section 6.5, de la répartition des capacités d'échantillonnage pour estimer un risque acarologique.

6.1 « Removal Sampling » : avantages et limites de la méthode adaptée aux tiques

L'application des méthodes d'échantillonnage par « removal sampling » aux tiques *I. ricinus* nous paraît être, à ce jour, une des méthodes les plus pertinentes pour le recueil de données nécessaires à l'estimation du taux d'échantillonnage τ et de la taille de la population des tiques à l'affût N_0 . Nous avons considéré la population des tiques à l'affût comme fermée et la probabilité de capture comme constante. L'échantillonnage par RS présente l'avantage d'être facile à mettre en œuvre (aucune technique particulière), généralisable à d'autres conditions d'échantillonnage (type de végétation, saison, ...). Cette méthode s'est révélée plus adaptée que les méthodes par Capture Marquage Recapture (CMR) pour lesquelles le nombre de recaptures a été faible et variable entre les unités d'observation. Contrairement aux CMR, les méthodes par RS ne génèrent pas de biais liés aux phases de capture et de marquage qui peuvent modifier le comportement des individus marqués (Crowcroft and Jeffreers, 1961; Gilbert et al., 2001).

Cependant, la méthode d'échantillonnage par RS est coûteuse en termes de temps et de main d'œuvre lorsque le nombre d'échantillonnages successifs à réaliser est élevé (Dodd and Dorazio, 2004). De plus, au cours des échantillonnages successifs, trois raisons principales peuvent générer des entrées ou des sorties dans la population à l'affût et ainsi remettre en question l'hypothèse forte de population fermée : (i) le milieu échantillonné se modifie au fur et à mesure des balayages successifs de la végétation par le drap et ce, d'autant plus que le type de végétation est mobile, comme des feuilles mortes par exemple. Cette modification du milieu donne alors accès à une nouvelle population de tiques située au niveau du sol et qui était, au départ des échantillonnages, inaccessible car couverte par la végétation ; (ii) les tiques actives au sol (en réhydratation ou à l'affût au sol) peuvent se mettre à l'affût sur le haut de la végétation pour répondre aux stimulus créés par les passages répétés du drap malgré l'intervalle de temps court entre deux échantillonnages successifs ; (iii) les tiques à l'affût qui n'ont pas été capturées par le drap bien que placées sur le haut de la végétation sont susceptibles de tomber de leur support suite aux vibrations générées par le passage du drap.

Si l'hypothèse de population fermée n'est pas vérifiée du fait de la mise à l'affût des tiques alors l'estimation du taux ne correspondra pas au taux d'échantillonnage

de la population à l'affût comme prévu initialement. Il correspondra au produit de deux probabilités, à savoir, la probabilité pour une tique d'être capturée par le drapeau lorsqu'elle est à l'affût (τ) et la probabilité pour une tique de se mettre à l'affût lorsque elle est active.

Le principal problème de l'approche par RS appliquée aux tiques est qu'il est très difficile d'estimer avec certitude la vraie valeur de la taille de la population de tiques à l'affût. Le RS est très utilisé pour estimer les abondances de poisson. Lorsque les estimations sont réalisées dans des milieux comme des lacs ou des rivières, il est envisageable exceptionnellement d'avoir accès à la vraie valeur de l'abondance afin de vérifier les estimations. Ceci doit bien évidemment se faire en considérant d'éthique. On peut, par exemple, assécher un lac pour capturer tous les poissons d'un pan de rivière. Cette proposition n'est pas applicable aux tiques à cause de la spécificité de leur comportement et du milieu dans lequel elles évoluent. On pourrait envisager d'extraire un mètre carré de végétation sur une profondeur de quelques centimètres pour le tamiser et compter ensuite le nombre réel de tiques présentes. Cependant, cette technique ne nous permettra pas de distinguer les tiques qui étaient à l'affût de celles qui sont actives au sol ou non actives. On pourrait également envisager de compter à la vue les tiques à l'affût, mais cette proposition est difficilement répétable sur de nombreuses unités d'observation et est plus ou moins facile à réaliser selon le type et la densité de la végétation. Malgré ces limites, à l'heure actuelle, la méthode par RS semble être la plus adaptée pour estimer la taille de la population des tiques. Un des moyens pour limiter les problèmes d'entrée et de sortie est de limiter le nombre d'échantillonnages successifs et de réduire au maximum l'intervalle de temps entre leurs échantillonnages successifs.

6.2 Valeurs estimées du taux d'échantillonnage des tiques *I. ricinus*

Les estimations du taux d'échantillonnage que nous avons obtenu (environ 50%) sont bien supérieures aux rares estimations de la littérature (inférieures à 10%) (Sonnenshine, 1994; Daniels et al., 2000; Talleklint-Eisen and Lane, 2000). L'unique étude qui utilise une approche de RS (Talleklint-Eisen and Lane, 2000) considère un délai entre les échantillonnages de l'ordre de la journée. Dans ce cas, l'hypothèse de population fermée est discutable car les tiques actives présentes au sol ont le temps d'immigrer dans la population à l'affût entre deux échantillonnages. De plus, les unités d'observations n'étant pas cloisonnées, certaines tiques à l'affût peuvent s'agripper à un hôte de passage et donc émigrer.

Dans notre étude, pour préserver cette hypothèse de population fermée, nous avons considéré un intervalle de temps entre deux échantillonnages successifs le plus

court possible, à savoir, 2 minutes et 30 secondes. Cet intervalle de temps comprend le temps nécessaire au passage du drap sur la végétation de l'unité d'observation (environ 30s) et l'inspection du drap pour le comptage des nymphes présentes. Lorsque le nombre de nymphes présentes sur le drap était trop important, cet intervalle de temps n'était parfois pas suffisant pour permettre une inspection complète du drap avant l'échantillonnage suivant ; dans cas, nous avons utilisé un autre drap (similaire en dimension et en matière).

Le fait de considérer un intervalle court entre deux échantillonnages présente deux avantages. D'abord, il permet de réaliser l'ensemble des échantillonnages successifs d'une même unité d'observation dans des conditions similaires (MacKenzie and Royle, 2005). Quand cet intervalle de temps est important alors les conditions d'échantillonnage (température, hygrométrie, vent, ...) vont varier d'un échantillonnage à l'autre ce qui peut engendrer, au fur et à mesure des échantillonnages successifs, (i) une modification de la taille de la population à l'affût *i.e.* en favorisant la mise à l'affût des tiques actives au sol ou en incitant les tiques à l'affût à descendre au sol pour se réhydrater ; (ii) une modification de la probabilité de capture en altérant la réactivité des tiques au passage du drap, en modifiant son efficacité (en cas d'humidité) ou en diminuant la hauteur de quête rendant les tiques inaccessibles ? ; dans ces cas, l'hypothèse de taux d'échantillonnage constant n'est plus vérifiée. Par conséquent, considérer un intervalle de temps long entre 2 échantillonnages successifs peut conduire à une confusion entre la variabilité du taux d'échantillonnage et la variabilité de la taille de la population à l'affût pour expliquer la variation du nombre de captures (Yoccoz et al., 2001; MacKenzie and Kendall, 2002). Le deuxième avantage du fait de considérer un intervalle court entre deux échantillonnages est qu'il permet de prendre en compte davantage d'échantillonnages successifs pour estimer le taux d'échantillonnage tout en préservant l'hypothèse de population fermée (dans la limite des contraintes du « removal sampling » discutées au paragraphe 6.1, *i.e.* modification du milieu et coûts d'échantillonnages). En effet, si on considère un intervalle de 10 minutes entre 2 échantillonnages, le 5^{ème} échantillonnage sera réalisé 40 minutes après le premier contre 15 minutes si l'intervalle est réduit à 2 minutes 30s. On peut ainsi supposer qu'en 15 minutes peu de tiques auront eu le temps d'immigrer dans la population à l'affût ; dans ce cas, l'hypothèse de population fermée est plus vraisemblable par rapport à 40 minutes d'échantillonnage. Cet avantage sera d'autant plus important que le taux d'échantillonnage sera faible. En effet, MacKenzie and Royle (2005) préconisent au moins 3 échantillonnages successifs pour un taux d'échantillonnage supérieur à 0,5 et davantage pour des taux inférieurs.

6.3 Importance de la variabilité du taux dans l'estimation de l'abondance

Estimer l'abondance de chaque unité d'observation en tenant compte du taux d'échantillonnage autorise la comparaison des estimations d'abondance entre différentes unités d'observation et la recherche des facteurs caractérisant les unités d'observation ayant un effet sur l'abondance. Plusieurs de nos résultats concordent pour souligner l'importance de prendre en compte la variabilité du taux d'échantillonnage dans l'estimation de l'abondance.

Premièrement, nous avons trouvé que les estimations du taux d'échantillonnage se situent à environ 50%. C'est exactement pour ce niveau de taux que la variabilité du nombre de capture est maximale. En effet, si X représente le nombre de tiques capturées avec une probabilité τ parmi une population à l'affût de taille N_0 alors X suit une loi binomiale de paramètres (N_0, τ) et sa variabilité $V(X) = N_0\tau(1 - \tau)$ est maximale pour $\tau = 0,5$ (cf. Fig. 6.1). Autrement dit, si on considère 2 unités d'observation dont les tailles de populations et le taux d'échantillonnage sont strictement égaux alors la variabilité des quantités de captures observées sur chacune des 2 unités d'observation sera maximale pour $\tau = 0,5$.

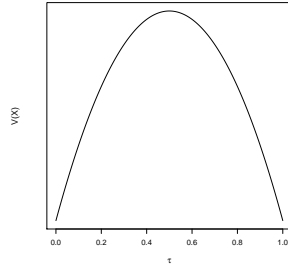


Fig. 6.1 : Variance d'une loi binomiale en fonction de la probabilité τ

Deuxièmement, bien que nos échantillonnages aient été réalisés dans des conditions environnementales plutôt homogènes, nous avons mis en évidence par deux approches complémentaires (recherche de facteurs de risque par des modèles glm d'abord (Chapitre 3) et estimation du taux d'échantillonnage et de la taille de population par un modèle hiérarchique bayésien ensuite (Chapitre 4)) que plusieurs facteurs jouaient un rôle significatif sur la variation du taux d'échantillonnage : le type de végétation (feuilles mortes ou arbustes) et l'heure d'échantillonnage (matin ou après-midi). Les conditions environnementales ont été plutôt homogènes car la météo a été stable au cours des campagnes d'échantillonnage, les unités d'observation ont été échantillonnées dans une forêt dont le couvert végétal était partout forestier et elles étaient situées dans une zone où population d'hôtes était importante.

Troisièmement, nous avons montré qu'en augmentant le nombre de d'échantillonnages successifs cela permettait de s'affranchir de l'influence de certains facteurs sur le taux d'échantillonnage. En effet, en considérant au moins 4 échantillonnages successifs, l'influence du type de végétation sur la probabilité pour une tique d'être capturée était gommée contrairement à l'heure d'échantillonnage qui semble toujours avoir une influence.

L'effet persistant de l'heure d'échantillonnage sur la probabilité de capture est vraisemblablement due à une différence de réactivité des tiques face au stimulus créé par le passage du drap selon l'heure d'échantillonnage. Cet effet est à mettre en relation avec la température et l'hygrométrie (i) qui modifient le niveau d'activité des tiques (Sonenshine, 1994) : le matin étant plus frais et plus humide (température moyenne à 14,1 °C et hygrométrie moyenne à 80,8%) que l'après midi (température moyenne à 17,9 °C et hygrométrie moyenne à 72,8%) ; (ii) et qui pourraient diminuer la hauteur de quête selon Randolph and Storey (1999).

Pour ce qui est de l'effet de la végétation, on peut supposer qu'il est lié à la probabilité de contact entre le drap et la population à l'affût qui est située sur le haut de la végétation (Milne, 1943; Ginsberg and Ewing, 1989; Schulze et al., 1997; Tack et al., 2011). En effet, en présence de feuilles mortes, la probabilité de contact entre le drap et la population à l'affût paraît plus importante qu'en présence d'arbustes. En multipliant les échantillonnages successifs, on augmente la probabilité de contact entre le drap et la population à l'affût ce qui permet de gommer les différences entre les types de végétation.

Il paraît donc crucial de prendre en compte le taux d'échantillonnage dans l'estimation d'abondance des unités d'observations. Cependant, il faut limiter le nombre d'échantillonnages successifs afin de limiter la modification du milieu et l'immigration de tiques dans la population à l'affût. La méthode d'échantillonnage par RS associé à la modélisation hiérarchique bayésienne permet d'atteindre cet objectif. En effet, cette modélisation hiérarchique nous a permis d'estimer simultanément l'abondance et le taux d'échantillonnage spécifique pour chaque unité d'observation en prenant en compte en tant qu'effet principal, le type de végétation et l'heure d'échantillonnage sur le taux d'échantillonnage ainsi qu'une fluctuation aléatoire liée à chacune des unités d'observation, ceci en ne considérant que peu d'échantillonnages successifs (entre 3 et 4). Cet avantage est non négligeable car il limite les coûts d'échantillonnages (temps, main d'œuvre) liés à la multiplication des échantillonnages successifs et préserve l'hypothèse de population fermée.

De plus, l'avantage de la modélisation hiérarchique est qu'elle permet de prendre en compte les processus à différentes échelles (site observé, parcelle, région, ...) et la structure de dépendance des observations (Royle and Dorazio, 2008; Soubeyrand and Chadoeuf, 2006) qui peuvent avoir une influence sur les données observées. Ces processus sont décrits au travers de paramètres observables (température, végétation)

ou non observables (fluctuation aléatoire du taux de détection entre unités d'observation de mêmes conditions par exemple) (Royle and Dorazio, 2008; Soubeyrand et al., 2007).

6.4 Choix des *a priori*

En l'absence d'information *a priori*, il est courant d'utiliser des lois non-informatives, *i.e.* pour lesquelles aucune valeur n'est privilégiée. Cependant, cette solution n'est pas envisageable dans le cas de données issues de « removal sampling » car dans ce cas, les estimateurs bayésiens divergent. La stabilisation des estimateurs passe donc par le choix d'*a priori* informatifs pour τ ou pour N_0 . Pour la modélisation bayésienne hiérarchique de nos données de removal sampling, nous avons considéré une loi *a priori* informative sur la taille de la population à savoir une loi uniforme sur $[1, 1000]$ car nous voulions surtout étudier sur les variabilités du taux d'échantillonnage. Concrètement, cela signifie que *a priori* la taille de la population à l'affût est une valeur comprise entre le nombre cumulé de capture et 1000.

Au vu de l'étude menée dans la dernière partie de la thèse, nous conseillons, en l'absence d'information *a priori*, de considérer une loi *a priori* vague (dflat) pour N_0 et une loi $\beta(3, 3)$ pour τ . Par contre, si l'on sait que τ est plutôt faible ($\tau = 0,10$), le choix de la loi *a priori* pour τ va se faire en fonction du paramètre (N_0 ou τ) pour lequel l'utilisateur souhaite avoir la meilleure estimation. S'il souhaite disposer d'une bonne estimation de τ alors il pourra opter en faveur d'une loi $\beta(2,1, 10)$. S'il souhaite disposer d'une bonne estimation de N_0 alors il pourra opter en faveur d'une loi $\beta(3, 3)$. Le taux d'échantillonnage étant un paramètre de nuisance sans intérêt en soi, on peut penser que l'utilisateur préférera disposer d'une bonne estimation de N_0 et donc utilisera une loi $\beta(3, 3)$ pour τ .

6.5 Estimation du risque acarologique, comment répartir la capacité d'échantillonnage ?

Nous venons de voir que la méthode par « removal sampling » permet d'estimer la taille de la population à l'affût à un instant donné, *i.e.* à l'instant de l'échantillonnage.

Deux perspectives majeures sont à considérer pour améliorer nos estimations par rapport au risque de se faire mordre par une tique dans l'environnement.

Tout d'abord, les acteurs (services prévention, utilisateurs, décideurs, ...) peuvent s'intéresser à estimer le risque représenté par l'ensemble des tiques susceptibles de se mettre à l'affût sur une unité d'observation. On pourrait alors envisager d'estimer, pour chaque unité d'observation, la taille de la population active qui se compose de la population à l'affût et de la population active au sol. Par conséquent, nous pourrions

envisager pour chaque unité d'observation de (1) réaliser entre 2 et 4 échantillonnages successifs en considérant un intervalle de temps court entre deux échantillonnages successifs afin d'estimer le taux d'échantillonnage et la taille de la population à l'affût et de (2) répéter cette expérience tous les x jours sur l'ensemble de la période d'activité des tiques afin de laisser le temps aux tiques actives au sol d'intégrer la population à l'affût. Pour préserver l'hypothèse selon laquelle la population soit fermée, il faudrait clôturer l'unité d'observation pour éviter le passage d'hôtes sur lesquels des tiques à l'affût pourraient s'agripper et émigrer. À partir de ces données, nous pourrions alors estimer la taille de la population active par la somme cumulée des captures journalières et décrire la dynamique de mise à l'affût de la population de tiques actives au cours de la période d'activité en comparant la taille de la population à l'affût journalière à la taille de la population active.

Deuxièmement, les acteurs (services prévention, utilisateurs, décideurs, ...) peuvent s'intéresser au niveau de risque associé à une zone qui est généralement bien plus grande qu'une unité d'observation (10m^2) par exemple, une forêt, une région, un paysage. La question est alors de savoir comment répartir l'effort d'échantillonnage dans la zone d'intérêt. Dans un contexte de ressources limitées en temps et en moyens humains, va se poser la question de la répartition de l'effort d'échantillonnage entre échantillonnages successifs et unités d'observation. Concrètement, va-t-on préférer échantillonner *une fois 30 unités d'observations* plutôt qu'échantillonner *trois fois 10 unités d'observations* ?

Un plan d'échantillonnage optimal serait un compromis entre la nécessité de réaliser plusieurs échantillonnages pour quelques unités d'observation pour estimer le taux d'échantillonnage et la nécessité d'explorer plusieurs unités d'observation pour représenter au mieux la distribution spatiale de la population. Royle et al. (2005) propose une stratégie générale qui consiste à échantillonner plus d'unités d'observation lorsque le niveau d'occupation du territoire est faible et de concentrer l'effort d'échantillonnage sur quelques unités d'observation lorsque le degré d'occupation est fort. Cependant, pour choisir entre ces deux propositions et donc déterminer le plan d'échantillonnage optimal, encore faut-il avoir une idée du modèle qui décrit la répartition spatiale des tiques. Cela revient à évaluer, à l'échelle de la zone d'intérêt, le niveau de variation du taux d'échantillonnage par rapport à celui de l'abondance de la population. En d'autres termes, à l'échelle de la zone d'intérêt, la variation du taux d'échantillonnage est-elle négligeable par rapport à celle de la taille de population des unités d'observation ? Toutes les études publiées montrent des variations importantes de N_0 . Notre travail a souligné, pour sa part, l'importance de la variabilité du taux d'échantillonnage. Cependant, nous n'avons pas assez de recul pour répondre à la question précédente.

Pour répondre à cette question, trois pistes peuvent être envisagées : (i) évaluer, à partir de simulations, notre capacité à mettre en évidence l'effet de facteurs sur

l'abondance lorsque le taux d'échantillonnage varie; (ii) proposer différents scénarios de répartition spatiale des tiques afin de déterminer, pour chacun des scénarios simulé, le plan d'échantillonnage optimal; (iii) inciter les responsables de projets à mettre en place des protocoles de collecte de tiques par « removal sampling » en considérant 3 échantillonnages successifs afin de pouvoir estimer la valeur du taux d'échantillonnage et sa variabilité dans différents contextes. L'accumulation de ces données nous donnera des indications empiriques sur la part relative de la variation du taux d'échantillonnage.

6.6 Conclusion

Au cours de ce travail de thèse, nous avons proposé une modélisation hiérarchique bayésienne pour estimer le taux d'échantillonnage et la taille de la population à partir de données de « removal sampling ». Cette modélisation s'est faite dans un contexte particulier où l'hypothèse de population fermée était fragilisée par la multiplication des échantillonnages successifs. De plus, dans un contexte où les distributions *a priori* non informatives sur les paramètres τ et N_0 conduisent à des estimateurs bayésiens non convergents, nous avons proposé de poser des lois *a priori* informatives $\beta(3;3)$ pour τ et dflat pour N_0 qui permettent de bonnes estimations du paramètre d'intérêt N_0 sans pour autant avoir trop d'influence sur les distributions *a posteriori*. La perspective majeure de ce travail est d'étudier comment répartir au mieux la force d'échantillonnage à l'échelle d'une parcelle *i.e.* étudier l'importance de la variabilité du taux d'échantillonnage par rapport à la variabilité de la répartition de la taille de la population des tiques.

Annexes

7.1 Pré-étude de CMR

7.1.1 Peu ou pas de recapture

Nous présentons les données de captures et de recaptures observées sur 4 unités d'observations pour lesquelles nous avons observé peu voire aucune recapture de nymphes marquées (cf. Fig. 7.1). Ce résultat est peut-être lié à un effet des phases de capture et de marquage sur la survie des nymphes marquées ou à un intervalle de temps trop court entre le lâcher et le 1^{er} échantillonnage qui n'a pas permis aux nymphes marquées de se remettre à l'affût.

7.1.2 Augmentation du nombre de recaptures au fur et à mesure des échantillonnages

Nous présentons les données de capture et de recaptures observées sur 3 unités d'observations pour lesquelles nous avons observé une augmentation progressive des quantités de captures de nymphes marquées au cours des premiers échantillonnages successifs (cf. Fig. 7.2). Ce résultat est peut-être lié à un intervalle de temps trop court entre le lâcher et le 1^{er} échantillonnage qui conduit à capturer les nymphes marquées au fur et à mesure qu'elles se mettent à l'affût.

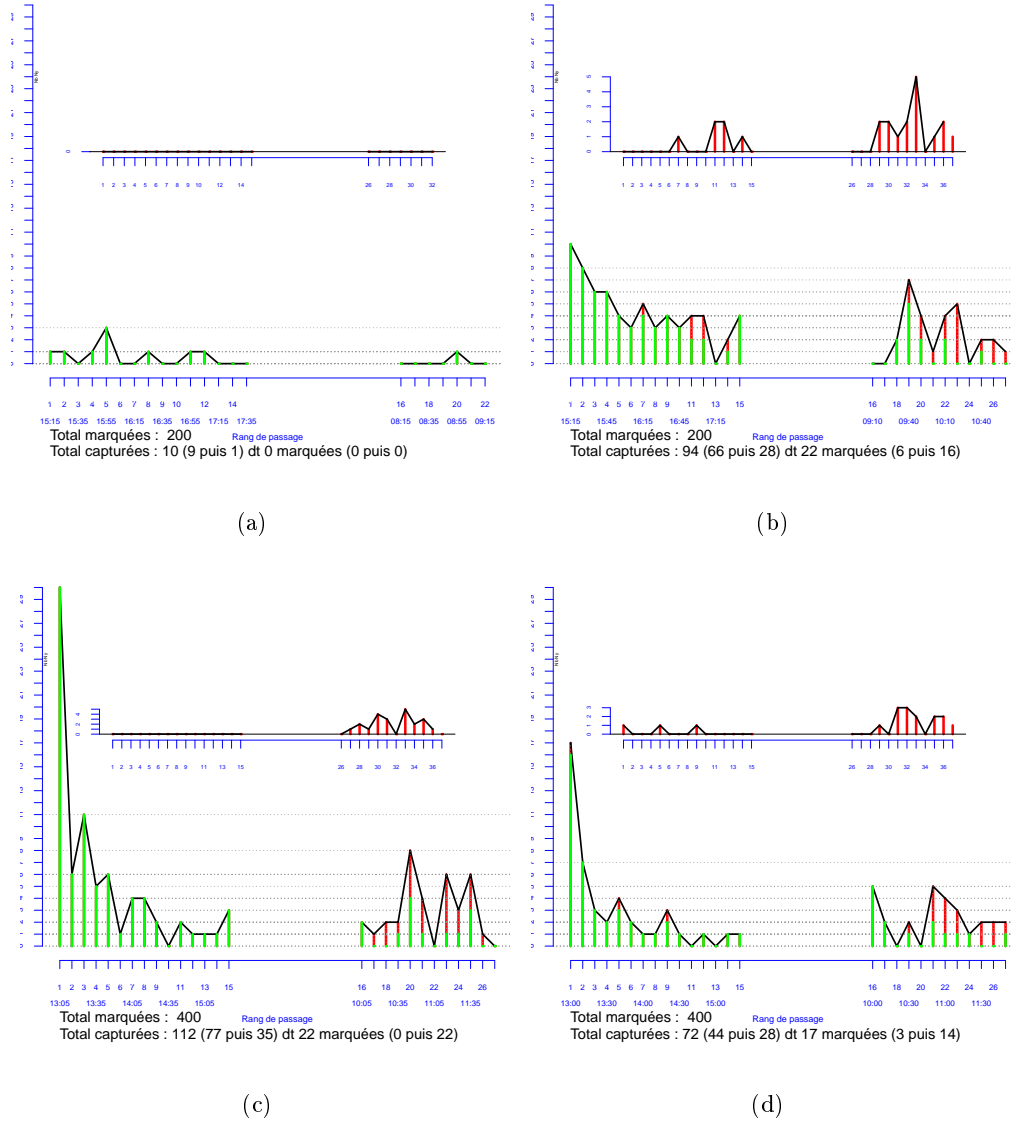


Fig. 7.1 : Exemple unités d'observations pour lesquelles nous avons eu très peu ou pas de recaptures. Nombre de captures de tiques marquées (en rouge) et non marquées (en vert) observées sur 4 unités d'observation. L'intervalle de temps entre deux échantillonnages successifs était de 10 minutes. Le 1^{er} échantillonnage a été réalisé 2h30 après le lâcher des nymphes marquées (vers 15h environ) et le 15^{ème} a été réalisé le lendemain matin (vers 9h environ). Type de végétation des unités d'observations (a), (b) et (c) : Feuilles mortes ; (d) : Feuilles mortes et arbustes.

7.1.3 Décroissance puis augmentation du nombre de recaptures au fur et à mesure des échantillonnages

Nous présentons les données de capture et de recaptures observées sur 3 unités d'observations pour lesquelles nous avons observé une augmentation progressive des

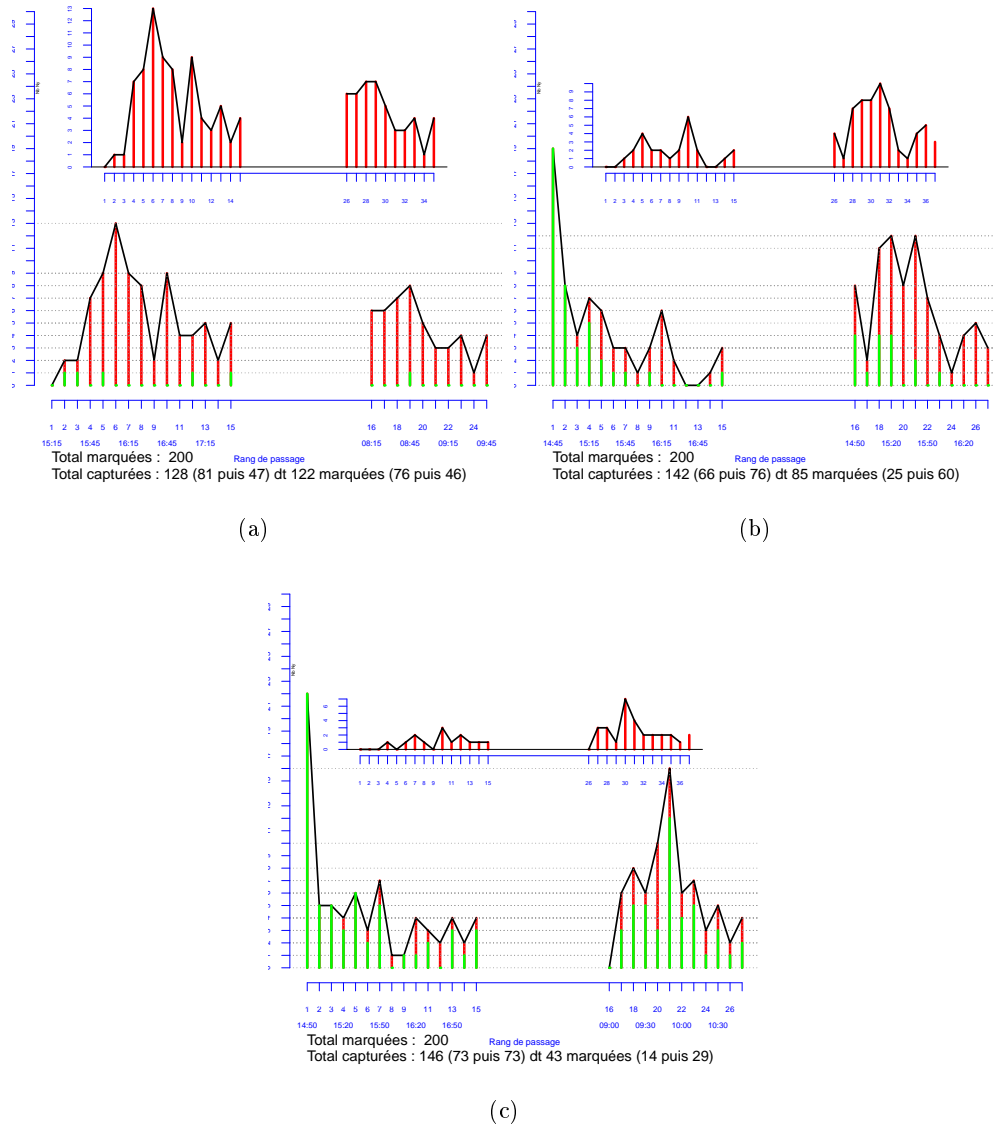


Fig. 7.2 : Exemple unités d'observations pour lesquelles nous avons observé une augmentation progressive des quantités de captures de nymphes marquées. Nombre de captures de tiques marquées (en rouge) et non marquées (en vert) observées sur 4 unités d'observation. L'intervalle de temps entre deux échantillonnages successifs était de 10 minutes. Le 1^{er} échantillonnage a été réalisé 2h30 après le lâcher des nymphes marquées (vers 15h environ) et le 15^{ème} a été réalisé le lendemain matin (vers 9h environ). Type de végétation des unités d'observations (a) et (b) : Feuilles mortes; (c) : Feuilles mortes et arbustes.

quantités de captures de nymphes marquées au cours des premiers échantillonnages successifs (cf. Fig. 7.3). Ce résultat est peut-être lié à une mise à l'affût des nymphes

actives au sol au fur et à mesure des échantillonnages successifs.

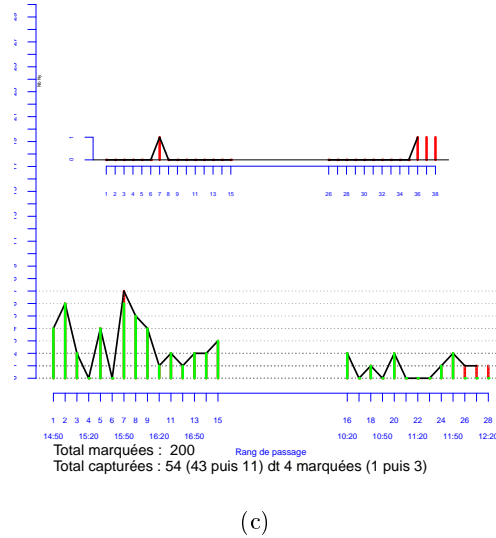
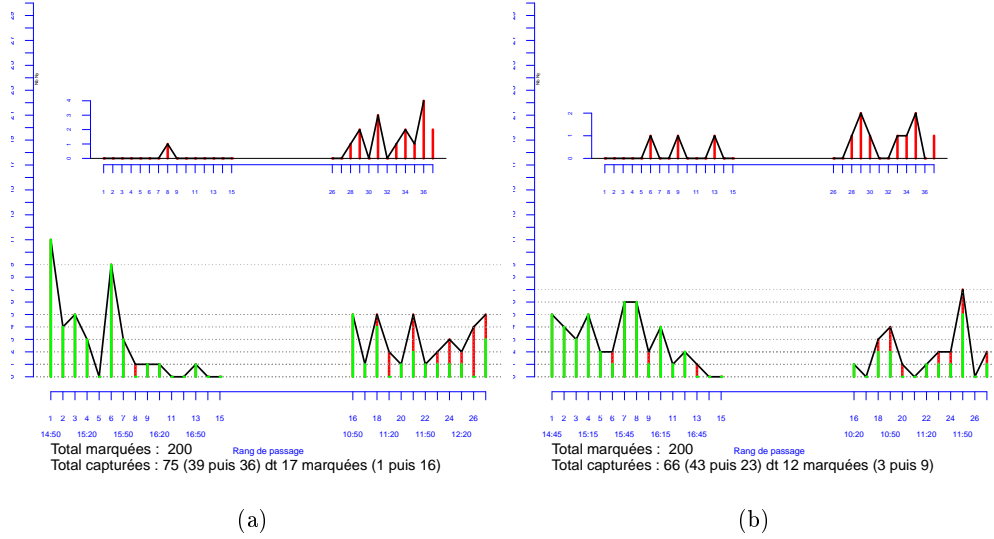


Fig. 7.3 : Exemple unités d'observations pour lesquelles nous avons observé une diminution suivie d'une augmentation des quantités de captures. Nombre de captures de tiques marquées (en rouge) et non marquées (en vert) observées sur 4 unités d'observation. L'intervalle de temps entre deux échantillonnages successifs était de 10 minutes. Le 1^{er} échantillonnage a été réalisé 2h30 après le lâcher des nymphes marquées (vers 15h environ) et le 15^{ème} a été réalisé le lendemain matin (vers 9h environ). Type de végétation des unités d'observation (a) : Feuilles mortes et arbustes; (b) et (c) : Herbe.

7.2 Vraisemblance et vraisemblance profilée

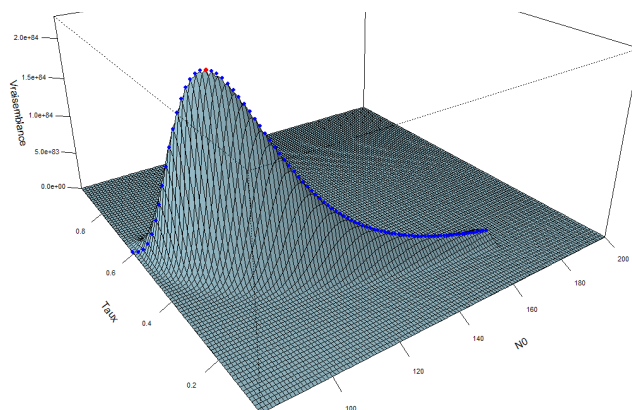


Fig. 7.4 : Vraisemblance et vraisemblance profilée (en bleu) pour la séquence simulée, $X = (43, 23, 15)$ avec $N_0 = 100$ et $\tau = 0.4$

Bibliographie

- Anderson, D. R. (2001). The need to get the basics right in wildlife field studies. *Wildlife Society Bulletin*, 29(4) :1294–1297.
- Anderson, D. R. (2003). Response to Engeman : Index values rarely constitute reliable information. *Wildlife Society Bulletin*, 31(1) :288–291.
- Bayes, M. and Price, M. (1763). An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. *Philosophical Transactions*, 53 :370–418.
- Bedrick, E. J. (1994). Maximum-likelihood estimation for the removal method. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 22(2) :285–293.
- Begon, M., Townsend, C. R., and Harper, J. L. (2009). *Ecology : from individuals to ecosystems*. John Wiley & Sons.
- Belozerov, V. N. (1982). Diapause and biological rhythms in ticks. In Obenchain, F. and Galun, R., editors, *Physiology of Ticks : Current Themes in Tropical Science Series*, volume 1, pages 469–500.
- Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*. Wiley, New York.
- Bohrmann, T. F., Christman, M. C., and Smith, S. J. (2012). Evaluating sampling efficiency in depletion surveys using hierarchical Bayes. *Canadian Journal of Fisheries and Aquatic Sciences*, 69(6) :1080–1090.
- Bolfarine, H., Leite, J. G., and Rodrigues, J. (1992). On the estimation of the size of a finite and closed population. *Biometrical Journal*, 34(5) :577–593.
- Borchers, D. L., Buckland, S. T., and Zucchini, W. (2002). *Estimating Animal Abundance : Closed Populations*. Statistics for Biology and Health. Springer, London.
- Bord, S., Druilhet, P., Gasqui, P., Abrial, D., and Vourc’h, G. (2014). Bayesian estimation of abundance based on removal sampling under weak assumption of closed

- population with catchability depending on environmental conditions. Application to tick abundance. *Ecological Modelling*, 274(0) :72–79.
- Boyard, C., Barnouin, J., Gasqui, P., and Vourc'h, G. (2007). Local environmental factors characterizing *Ixodes ricinus* nymph abundance in grazed permanent pastures for cattle. *Parasitology*, 134 :987–994.
- Boyard, C., Vourc'h, G., and Barnouin, J. (2008). The relationships between *Ixodes ricinus* and small mammal species at the woodland-pasture interface. *Experimental and Applied Acarology*, 44(1) :61–76.
- Brun, M., Abraham, C., Jarry, M., Dumas, J., Lange, F., and Prevost, E. (2011). Estimating an homogeneous series of a population abundance indicator despite changes in data collection procedure : A hierarchical Bayesian modelling approach. *Ecological Modelling*, 222(5) :1069–1079.
- Buckland, S. T., Anderson, D. R., Burnham, K. P., and Laake, J. L. (2005). Distance Sampling. In *Encyclopedia of Biostatistics*. John Wiley & Sons, Ltd.
- Burnham, K. P. (1981). Summarizing remarks : Environmental influences. In *Estimating Numbers of terrestrial birds (CJ Ralph and JM Scott, Eds.)*. *Studies in Avian Biology*, (6), pages 324–325.
- Carle, F. L. and Strub, M. R. (1978). A new method for estimating population size from removal data. *Biometrics*, 34(4) :621–630.
- Conroy, M. J. (1996). Abundance indices. In *Measuring and monitoring biological diversity : Standard methods for mammals*, pages 179–192. Smithsonian Institution Press.
- Coreau, A., Pinay, G., Thompson, J. D., Cheptou, P.-O., and Mermet, L. (2009). The rise of research on futures in ecology : rebalancing scenarios and predictions. *Ecology Letters*, 12(12) :1277–1286.
- Crowcroft, P. and Jeffreers, J. N. R. (1961). Variability in the behaviour of wild house mice (*Mus Musculus* L.) towards live traps. *Proceedings of the Zoological Society of London*, 137(4) :573–582.
- Daniels, T. J., Falco, R. C., and Fish, D. (2000). Estimating population size and drag sampling efficiency for the blacklegged tick (Acari : Ixodidae). *Journal of Medical Entomology*, 37(3) :357–363.
- Dauphin, G., Prevost, E., Adams, C. E., and Boylan, P. (2009). A Bayesian approach to estimating Atlantic salmon fry densities using a rapid sampling technique. *Fisheries Management and Ecology*, 16(5) :399–408.

- Dodd, C. K. and Dorazio, R. M. (2004). Using counts to simultaneously estimate abundance and detection probabilities in a salamander community. *Herpetologica*, 60(4) :468–478.
- Ellison, A. M. (2004). Bayesian inference in ecology. *Ecology Letters*, 7(6) :509–520.
- Elphick, C. S. (2008). How you count counts : the importance of methods research in applied ecology. *Journal of Applied Ecology*, 45(5) :1313–1320.
- Estrada-Peña, A., Gray, J. S., Kahl, O., Lane, R. S., and Nijhof, A. M. (2013). Research on the ecology of ticks and tick-borne pathogens – methodological principles and caveats. *Frontiers in cellular and infection microbiology*, 3.
- Estrada-Pena, A., Martinez, J. M., Sanchez Acedo, C., Quilez, J., and Del Cacho, E. (2004). Phenology of the tick, *Ixodes ricinus*, in its southern distribution range (central Spain). *Medical and Veterinary Entomology*, 18(4) :387–397.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis*. Chapman & Hall/CRC.
- Gigon, F. (1985). *Biologie d’Ixodes ricinus sur le Plateau suisse*. PhD thesis, Neuchâtel.
- Gilbert, D., McKenzie, J., and Davies, N. (2001). Evidence from tag recapture experiments that fish learn to avoid fishing gear. *Journal of Agricultural, Biological, and Environmental Statistics*, 6(2) :281–291.
- Ginsberg, H. S. and Ewing, C. P. (1989). Comparison of flagging, walking, trapping, and collecting from hosts as sampling methods for northern deer ticks, *Ixodes dammini*, and lone-star ticks, *Amblyomma americanum*; (Acari : Ixodidae). *Experimental and Applied Acarology*, 7(4) :313–322.
- Gove, J. H., Linder, E., and Tzilkowski, W. M. (1995). Biomodality of the combined removal and signs-of-activities estimator for sampling closed animal populations. *Environmental and Ecological Statistics*, 3(1) :65–78.
- Gray, J. S. (1982). The development and questing activity of *Ixodes ricinus* (L.) (Acari : Ixodidae) under field conditions in Ireland. *Bulletin of Entomological Research*, 72(2) :263–270.
- Gray, J. S. (1985). The development and seasonal activity of the tick of *Ixodes ricinus* : a vector of Lyme borreliosis. *Review of medical and veterinary entomology*, (79) :323–333.

- Gray, J. S., Turley, T., and Strickland, K. L. (1978). Studies on ecology of sheep tick, *Ixodes ricinus* in Co Wicklow, Ireland. *Irish Veterinary Journal*, 32(2) :25–34.
- Gu, W. D. and Swihart, R. K. (2004). Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. *Biological Conservation*, 116(2) :195–203.
- Guégan, J.-F., Renaud, F., and Chapitre, V. (2004). *Vers une écologie de la sant  *, chapter 5, pages 100–135.   ditions adpf.
- Haldane, J. B. S. (1932). A note on inverse probability. *Mathematical Proceedings of the Cambridge Philosophical Society*, 28(01) :55–61.
- Harding, C. M., Heathwood, A. W., Hunt, R. G., and Read, K. L. Q. (1984). The estimation of animal population size by the removal method. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 33(2) :196–202.
- Hayne, D. W. (1949). An examination of the strip census method for estimating animal populations. *The Journal of Wildlife Management*, 13(2) :pp. 145–157.
- Hirst, D. (1994). An improved removal method for estimating animal abundance. *Biometrics*, 50(2) :501–505.
- Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 186(1007) :453–461.
- Jennelle, C. S., Cooch, E. G., Conroy, M. J., and Senar, J. C. (2007). State-specific detection probabilities and disease prevalence. *Ecological Applications*, 17(1) :154–167.
- Johnson, D. H. (1995). Point counts of birds : what are we estimating ? In *Monitoring bird populations by point counts*, pages 117–123.
- Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., and Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature*, 451(7181) :990–993.
- Kahl, O. and Kn  lle, W. (1988). Water vapour uptake from subsaturated atmospheres by engorged immature ixodid ticks. *Experimental and Applied Acarology*, 4(1) :73–83.
- Laplace, P. S. (1786). Sur les naissances, les mariages et les morts histoire de l’Academie Royale des Sciences.

- Lebreton, J.-D., Burnham, K. P., Clobert, J., and Anderson, D. R. (1992). Modeling survival and testing biological hypotheses using marked animals : a unified approach with case studies. *Ecological monographs*, 62(1) :67–118.
- Lees, A. D. (1946). The water balance in *Ixodes ricinus* L. and certain other species of ticks. *Parasitology*, 37(1-2) :1–20.
- Lees, A. D. and Milne, A. (1951). The seasonal and diurnal activities of individual sheep ticks (*Ixodes ricinus* L.). *Parasitology*, 41(3-4) :189–208.
- Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*. Springer Texts in Statistics. Springer.
- Leslie, P. H. and Davis, D. H. S. (1939). An attempt to determine the absolute number of rats on a given area. *Journal of Animal Ecology*, 8(1) :94–113.
- Lindén, A. and Mäntyniemi, S. (2011). Using the negative binomial distribution to model overdispersion in ecological count data. *Ecology*, 92(7) :1414–1421.
- MacKenzie, D. I. and Kendall, W. L. (2002). How should detection probability be incorporated into estimates of relative abundance ? *Ecology*, 83(9) :2387–2393.
- MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., and Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83(8) :2248–2255.
- MacKenzie, D. I. and Royle, J. A. (2005). Designing occupancy studies : general advice and allocating survey effort. *Journal of Applied Ecology*, 42(6) :1105–1114.
- MacLeod, J. (1932). The bionomics of *Ixodes ricinus* L, the "sheep tick" of Scotland. *Parasitology, Cambridge*, Vol. 24(No. 3) :pp. 382–400.
- Mantyniemi, S., Romakkaniemi, A., and Arjas, E. (2005). Bayesian removal estimation of a population size under unequal catchability. *Canadian Journal of Fisheries and Aquatic Sciences*, 62(2) :291–300.
- Marsot, M., Chapuis, J.-L., Gasqui, P., Dozières, A., Masségli, S., Pisanu, B., Ferquel, E., and Vourc'h, G. (2013). Introduced siberian chipmunks *Tamias sibiricus barberi* contribute more to Lyme Borreliosis risk than native reservoir rodents. *PLoS ONE*, 8(1) :e55377.
- Matuschka, F. R. and Spielman, A. (1986). The emergence of Lyme disease in a changing environment in North-America and Central-Europe. *Experimental and Applied Acarology*, 2(4) :337–353.

- McClintock, B. T., Nichols, J. D., Bailey, L. L., MacKenzie, D. I., Kendall, W. L., and Franklin, A. B. (2010). Seeking a second opinion : uncertainty in disease ecology. *Ecology Letters*, 13(6) :659–674.
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability.
- Milne, A. (1943). The comparison of sheep-tick populations (*Ixodes ricinus* L.). *Annals of Applied Biology*, 30 :240–250.
- Moran, P. A. P. (1951). A Mathematical Theory of Animal Trapping. *Biometrika*, 38(3/4) :307–311.
- Morse, S. S. (1995). Factors in the emergence of infectious diseases. *Emerging infectious diseases*, 1(1) :7.
- Nefedova, V. V., Korenberg, E. I., Gorelova, N. B., and Kovalevskii, Y. V. (2004). Studies on the transovarial transmission of *Borrelia burgdorferi sensu lato* in the taiga tick *Ixodes persulcatus*. *Folia Parasitologica*, 51(1) :67–71.
- Ostfeld, R. S., Hazler, K. R., and Cepeda, O. M. (1996). Temporal and spatial dynamics of *Ixodes scapularis* (Acari : Ixodidae) in a rural landscape. *Journal of Medical Entomology*, 33(1) :90–95.
- Parola, P. (2004). Tick-borne rickettsial diseases : emerging risks in Europe. *Comparative Immunology Microbiology and Infectious Diseases*, 27(5) :297–304.
- Parola, P. and Raoult, D. (2001). Ticks and tickborne bacterial diseases in humans : an emerging infectious threat.
- Perret, J.-L., Guerin, P. M., Diehl, P. A., Vlimant, M., and Gern, L. (2003). Darkness induces mobility, and saturation deficit limits questing duration, in the tick *Ixodes ricinus*. *Journal of Experimental Biology*, 206(11) :1809–1815.
- Pollock, K. H., Nichols, J. D., Simons, T. R., Farnsworth, G. L., Bailey, L. L., and Sauer, J. R. (2002). Large scale wildlife monitoring studies : statistical methods for design and analysis. *Environmetrics*, 13(2) :105–119.
- Randolph, S. E. (2004). Tick ecology : processes and patterns behind the epidemiological risk posed by ixodid ticks as vectors. *Parasitology*, 129 :S37–S65.
- Randolph, S. E. and Craine, N. G. (1995). General framework for comparative quantitative studies on transmission of tick-borne diseases using Lyme borreliosis in Europe as an example. *Journal of Medical Entomology*, 32(6) :765–777.

- Randolph, S. E., Green, R. M., Hoodless, A. N., and Peacey, M. F. (2002). An empirical quantitative framework for the seasonal population dynamics of the tick *Ixodes ricinus*. *International Journal for Parasitology*, 32(8) :979–989.
- Randolph, S. E. and Storey, K. (1999). Impact of microclimate on immature tick-rodent host interactions (Acari : Ixodidae) : Implications for parasite transmission. *Journal of Medical Entomology*, 36(6) :741–748.
- Rauter, C. and Hartung, T. (2005). Prevalence of *Borrelia burgdorferi sensu lato* genospecies in *Ixodes ricinus* ticks in Europe : a metaanalysis. *Applied and environmental microbiology*, 71(11) :7203–7216.
- Rivot, E., Prevost, E., Cuzol, A., Bagliniere, J.-L., and Parent, E. (2008). Hierarchical Bayesian modelling with habitat and time covariates for estimating riverine fish population size by successive removal method. *Canadian Journal of Fisheries and Aquatic Sciences*, 65(1) :117–133.
- Rosenstock, S. S., Anderson, D. R., Giesen, K. M., Leukering, T., and Carter, M. F. (2002). Landbird counting techniques : Current practices and an alternative. *Auk*, 119(1) :46–53.
- Royle, J. A. (2004a). Modeling Abundance Index Data from Anuran Calling Surveys Model. *Conservation Biology*, 18(5) :1378–1385.
- Royle, J. A. (2004b). N-Mixture models for estimating population size from spatially replicated counts. *Biometrics*, 60(1) :108–115.
- Royle, J. A. (2006). Site occupancy models with heterogeneous detection probabilities. *Biometrics*, 62(1) :97–102.
- Royle, J. A. and Dorazio, R. M. (2006). Hierarchical models of animal abundance and occurrence. *Journal of Agricultural, Biological, and Environmental Statistics*, 11(3) :249–263.
- Royle, J. A. and Dorazio, R. M. (2008). *Hierarchical modeling and inference in ecology : the analysis of data from populations, metapopulations and communities*. Academic Press.
- Royle, J. A., Nichols, J. D., and Kéry, M. (2005). Modelling occurrence and abundance of species when detection is imperfect. *Oikos*, 110(2) :353–359.
- Schmidt, B. R. (2005). Monitoring the distribution of pond-breeding amphibians when species are detected imperfectly. *Aquatic Conservation : Marine and Freshwater Ecosystems*, 15(6) :681–692.

- Schnute, J. (1983). A new approach to estimating populations by the removal method. *Canadian Journal of Fisheries and Aquatic Sciences*, 40(12) :2153–2169.
- Schoeler, G. B. and Lane, R. S. (1993). Efficiency of transovarial transmission of the Lyme-disease Spirochete, *Borrelia Burgdorferi*, in the Western Blacklegged Tick, *Ixodes Pacificus* (Acari, Ixodidae). *Journal of Medical Entomology*, 30(1) :80–86.
- Schulze, T. L., Jordan, R. A., and Hung, R. W. (1997). Biases associated with several sampling methods used to estimate abundance of *Ixodes scapularis* and *Amblyomma americanum* (Acari : Ixodidae). *Journal of Medical Entomology*, 34(6) :615–623.
- Schwarz, C. J. and Seber, G. A. F. (1999). Estimating animal abundance : Review III. *Statistical Science*, 14(4) :427–456.
- Seber, G. A. F. (1982). The estimation of animal abundance : Griffin.
- Sonenshine, D. E. (1994). *Ecological Dynamics of tick-borne zoonoses*. Oxford.
- Soubeyrand, S. and Chadoeuf, J. (2006). Residual-based specification of a hidden random field included in a hierarchical spatial model. *Computational Statistics and Data Analysis*, 51(12) :6404–6422.
- Soubeyrand, S., Thebaud, G., and Chadoeuf, J. (2007). Accounting for biological variability and sampling scale : a multi-scale approach to building epidemic models. *Journal of the Royal Society Interface*, 4 :985–997.
- Tack, W., Madder, M., De Frenne, P., Vanhellemont, M., Gruwez, R., and Verheyen, K. (2011). The effects of sampling method and vegetation type on the estimated abundance of *Ixodes ricinus* ticks in forests. *Experimental and Applied Acarology*, 54(3) :285–292.
- Talleklint-Eisen, L. and Lane, R. S. (2000). Spatial and temporal variation in the density of *Ixodes pacificus* (Acari : Ixodidae) nymphs. *Environmental Entomology*, 29(2) :272–280.
- Thompson, W. L., White, G. C., and Gowan, C. (1998). *Monitoring Vertebrate Populations*. Elsevier Science.
- Truong-Van, B. and Prouzet, P. (2001). Méthodes de captures successives : synthèse, quelques contributions, une application. *Journal de la Société Française de Statistiques*, 142(3).
- Tylianakis, J. M., Didham, R. K., Bascompte, J., and Wardle, D. A. (2008). Global change and species interactions in terrestrial ecosystems. *Ecology Letters*, 11(12) :1351–1363.

- Tyre, A. J., Tenhumberg, B., Field, S. A., Niejalke, D., Parris, K., and Possingham, H. P. (2003). Improving precision and reducing bias in biological surveys : Estimating false-negative error rates. *Ecological Applications*, 13(6) :1790–1801.
- Walker, A. R. (2001). Age structure of a population of *Ixodes ricinus* (Acari : Ixodidae) in relation to its seasonal questing. *Bulletin of Entomological Research*, 91(1) :69–78.
- White, G. C., Leffler, B., and Laboratory, L. A. N. (1982). *Capture-recapture and removal methods for sampling closed populations*. LA-8787-NERP. Los Alamos National Laboratory.
- Williams, B. K., Nichols, J. D., and Conroy, M. J. (2002). *Analysis and Management of Animal Populations*. Academic Press, San Diego, USA & London, UK, 1st edition.
- Wyatt, R. J. (2002). Estimating riverine fish population size from single- and multiple-pass removal sampling using a hierarchical model. *Canadian Journal of Fisheries and Aquatic Sciences*, 59(4) :695–706.
- Yoccoz, N. G., Nichols, J. D., and Boulinier, T. (2001). Monitoring of biological diversity in space and time. *Trends in Ecology and Evolution*, 16(8) :446–453.
- Zippin, C. (1956). An evaluation of the removal method of estimating animal populations. *Biometrics*, 12(2) :163–189.
- Zippin, C. (1958). The removal method of population estimation. *The Journal of Wildlife Management*, 22(1) :pp. 82–90.

Cette thèse a été partiellement financée par le projet européen FP7-261504 EDENext et est répertoriée par le comité de direction d'EDENext en tant que EDENext312 (<http://www.edenext.eu>). Le contenu de cette publication relève de la seule responsabilité de l'auteur et ne reflète pas nécessairement les opinions de la Commission Européenne.

This study was partially funded by EU grant FP7-261504 EDENext and is catalogued by the EDENext Steering Committee as EDENext312 (<http://www.edenext.eu>). The contents of this publication are the sole responsibility of the authors and don't necessarily reflect the views of the European Commission.

Résumé

L'estimation des abondances de population est essentielle pour comprendre les dynamiques de population, les interactions entre espèces et estimer les risque de transmission d'agents pathogènes dans les populations. Plusieurs méthodes d'échantillonnages, basées sur des hypothèses spécifiques permettent d'estimer ces abondances : les méthodes par comptages uniques, par « distance sampling », par échantillonnages successifs ou par capture marquage recapture. Nous nous sommes intéressés à l'abondance des tiques *Ixodes ricinus*, vecteurs de nombreux agents pathogènes. Cette abondance est classiquement estimée par le nombre de tiques capturées lors d'échantillonnages uniques réalisés sur différentes unités d'observation. L'hypothèse forte de cette méthode est que le taux d'échantillonnage est constant entre les unités d'observation. Cependant, de nombreuses études remettent en cause cette hypothèse forte et suggèrent que le taux d'échantillonnage est variable selon les conditions d'échantillonnage (type de végétation, ...) mais ne prennent pas en compte ce taux d'échantillonnage pour autant.

A partir d'une méthode d'échantillonnage par « removal sampling » (RS), (i) nous avons montré que les conditions environnementales influençaient le taux d'échantillonnage et l'indicateur d'abondance usuel *i.e.* le nombre de tiques capturées lors d'un seul échantillonnage (ii) nous avons proposé une méthode pour détecter l'indicateur d'abondance, basés sur le nombre cumulé de capture, le moins soumis aux variations du taux ; (iii) par une approche Bayésienne hiérarchique, nous avons estimé simultanément l'abondance de tiques des unités d'observation et la valeur du taux d'échantillonnage en fonction du type de végétation et de l'heure d'échantillonnage. Nous avons montré que le taux d'échantillonnage sur des arbustes (entre 33.9% et 47.4%) était significativement inférieur au taux d'échantillonnage sur des feuilles mortes (entre 53.6% et 66.7%).

De plus, nous avons montré que le modèle RS tend vers un modèle de Poisson iid lorsque la taille de la population N_0 tends vers l'infini ce qui pose des problèmes d'indétermination pour estimer les paramètres N_0 et τ , le taux d'échantillonnage. Nous avons également montré que (i) les estimateurs Bayésiens divergent lorsque les lois *a priori* sont des lois vagues ; (ii) les lois *a priori* $\beta(a, b)$ avec $a > 2$ sur τ conduisaient à des estimateurs Bayésiens convergents. Enfin, nous avons proposé des recommandations quant au choix des lois *a priori* pour τ afin d'obtenir de bonnes estimations pour N_0 ou pour τ .

Nous discutons de la pertinence des méthodes RS pour les tiques et des perspectives envisageables pour (i) estimer le risque acarologique représenté par la population de tiques potentiellement actives sur une unité d'observation, (ii) estimer un risque à l'échelle d'une parcelle, à savoir comment répartir l'effort d'échantillonnage entre le nombre d'unités d'observation et le nombre d'échantillonnages successifs par unités d'observation.

Mots-clés : Abondance, Taux d'échantillonnage, Removal sampling, Approche Hiérarchique Bayésienne, distribution *a priori*

Abstract

The estimation of animal abundance is essential to understand population dynamics, species interactions and disease patterns in populations and to estimate the risk of pathogens transmission. Several sampling methods such as single counts, distance sampling, removal sampling or capture mark recapture could be used to estimate abundance. In this study, we are investigated the abundance of *Ixodes ricinus* ticks, which are involved in the transmission of many pathogens. Tick abundance is commonly estimated by the number of nymphs captured during a single observation (a cloth dragged on a given surface). In this case, analyses of abundance patterns assumes that the probability of detecting a tick, hence the sampling rate, remains constant across the observations. In practice, however, this assumption is often not satisfied as the sampling rate may fluctuate between observation plots. The variation of sampling rate is never taken into account in estimations of tick abundance.

Using a removal sampling design (RS), (i) we showed that the sampling rate and the usual abundance indicator (based on a single drag observation per spot) were both influenced by environmental conditions ; (ii) we proposed a method to determine the abundance indicator the least influenced by sampling rate variations ; (iii) using a hierarchical Bayesian model, we estimated simultaneously the abundance and the sampling rate according the type of vegetation and the time of sampling. The sampling rate varied between 33.9% and 47.4% for shrubs and 53.6% and 66.7% for dead leaves.

In addition, we show that the RS model tends to Poisson iid model when the population size N_0 tends to infinite. This result conduct to infinite estimations for N_0 . We show that (i) Bayesian estimators were divergent for vague prior ; (ii) $\beta(a, b)$ prior for $a > 2$ on τ conduct to convergent estimators. Then, we proposed recommendations for prior choice for τ parameter to give good estimations of N_0 or τ .

We discuss the relevance of RS for ticks and the possible perspectives to (i) estimate the acarologic risk associated to all potential active ticks for a given spot, (ii) estimate the risk at the larger scale, *i.e.* how to distribute the sampling effort between number of spot and number of consecutive sampling by spot.

Key-words : Abundance, Sampling rate, Removal sampling, Hierarchical Bayesian approach, prior distribution